Evaluating the Message:
The relationship between compliance rate
and the subject of a practice guideline

Roberto Grilli M.D., Jonathan Lomas M.A.

Centre for Health Economics and Policy Analysis
and
Department of Clinical Epidemiology and Biostatistics
McMaster University

1 Dr. Grilli is now located at the Istituto di Ricerche Farmacologiche "Mario Negri", Milano (Italy).

Running head: compliance with practice guidelines

Key words: review, practice guidelines, quality of care

Cite as: Grilli, R. and Lomas, J. "Evaluating the Message: The relationship between compliance rate and the subject of a practice guideline", McMaster University Centre for Health Economics and Policy Analysis
ACKNOWLEDGEMENTS

We would like to thank Brian Haynes, Paul Mazmanian, Andy Oxman and Chris Woodward for comments on an earlier draft of this paper. Support for Dr. Grilli was provided by the Associazione Italiana per la Ricerca sul Cancro and Prof. Lomas receives Career Scientist funding from the Ontario Ministry of Health.
ABSTRACT

Objective: To explore the relationship between providers' compliance and some key aspects of the clinical messages in practice guidelines.

Data Source: Studies published in the English language medical literature between 1980 and 1991 retrieved through MEDLINE and relevant review articles in the field.

Study Selection: All published studies providing compliance rates with practice guidelines and endorsed by official organizations were eligible for the study.

Data Extraction: The clinical content and the reported compliance rate were gathered for each recommendation in the 23 studies selected. The medical and surgical procedures addressed by 143 recommendations were identified according to specialty area, type of procedure (diagnostic, surgical, etc.) and were independently classified by the authors as being high or low on complexity, trialability and observability.

Data Synthesis: The mean compliance rate with the 143 clinical recommendations was 54.5% (95%CI: 50.2% - 58.9%), with those in the specialty areas of cardiology and oncology showing the highest compliance (mean 63.6% and 62.2%, respectively). Recommendations concerning procedures with high complexity had lower compliance rates than those low on complexity (41.9% vs 55.9%; p=0.05), and those judged to be high on trialability had higher compliance rates than those low on trialability (55.6% vs 36.8%; p=0.03). Overall, all the considered characteristics of the clinical recommendations in the practice guidelines could account for no more than 47% of the observed variability in compliance rates.

Conclusions: The target area of practice and the complexity and trialability of the recommended procedure appear to be useful, if partial, predictors of the level of compliance with a practice guideline.
INTRODUCTION

Practice guidelines have become increasingly popular over the past ten years (1). Health care research continues to produce important improvements in the diagnosis and treatment of patients. Medical organizations have shown a growing interest in the establishment and dissemination of clinical policies for different practice areas in order to improve quality of care. Government agencies have seen practice guidelines as a potentially useful tool in promoting a more cost-effective use of resources in health care, and in reducing variations in practice styles (2,3).

Even though the goals of the existing recommendations for clinical practice vary somewhat across the sponsoring organizations, they are based on the assumption that synthesizing and conveying the available scientific knowledge as a message to physicians through a credible source and in an easily understandable form can, by itself, improve providers' performance (4).

This optimistic view of the impact of practice guidelines on quality of care has not, however, been fully supported by the available research studies. It appears that most practice guidelines are only partly implemented even after active attempts to change physicians' behaviour. Some effort has gone into documenting the extra-clinical influences on physician decision-making that might explain these low compliance rates with practice guidelines (5,6). Much less effort, however, has gone into relating observed compliance rates to aspects of the specific clinical message of a practice guideline. Are recommendations, for instance, in one area of practice or for one type of procedure more or less likely to be implemented?

Drawing on the marketing literature Winkler et al. isolated "the message", and its content, as a potentially important determinant of a guideline's acceptability (7). (Their other determinants were the source, the channel, the audience, and the setting.)
Furthermore, the diffusion of innovations literature suggests that certain characteristics of a recommended procedure - such as its compatibility with a practitioner's existing activities or its degree of complexity - would influence compliance (8-12).

Practice guidelines are one way of 'packaging' a message for clinicians and the subject of such a message is therefore a potentially important influence on potential compliance with a particular practice guideline (13, 14). On the basis of the existing theoretic literature and our own hypotheses we, therefore, identified a number of possible aspects of the subject of the message that might influence compliance: the type of procedure involved (e.g. diagnostic versus therapeutic); the area of clinical practice (e.g. referred versus primary care); and, following Rogers (8), the characteristics of the guideline's recommended approach (e.g. its complexity, trialability and observability).

To assess the impact of these factors, we identified practice guideline studies that included a measure of compliance with recommendations. This allowed us to explore the relationship between the subject of a practice guideline's "message" and the extent of its acceptance in clinical practice. Also, by adopting the assumption that the guidelines represented good quality of care according to available scientific knowledge, the compliance rates may be taken as a proxy for the overall quality of care across different practice areas.
MATERIALS AND METHODS

The official recommendations in practice guidelines were the unit of analysis of the study. They were identified through papers published in the English language medical literature from 1980 to 1991 which provided the extent of compliance of clinical practice with the official recommendations in the guideline. Studies were identified searching through MEDLINE (search terms:standards, guidelines, official policy, consensus development, evaluation studies), the reference lists of relevant reviews in the field (1,15), published bibliographies (16) and personal contacts.

To be eligible for the study papers had to present compliance rates with practice guidelines developed by official organizations (medical profession organizations, Task Forces, government agencies), and had to target providers as the audience. Within each study we defined as "recommendation" every statement (as reported in the text or in tables) attributed to one of the above described official organizations and describing how a physician should manage a specific clinical situation. Recommendations from studies assessing the impact of practice guidelines developed at some local level (i.e. hospital, department, community) were not included, neither were those from clinical trials comparing different implementation strategies.

Compliance rates were extracted from the studies according to either of two definitions. For those studies using physicians as their unit of analysis (i.e. studies using physician self-report as the data source), it was defined as the proportion of providers acting according to those recommendations. For those studies using cases as the unit of analysis (i.e. studies using chart audit or administrative data as the source), it was defined as the proportion of patients treated according to the recommendations.

Overall, 47 studies were identified and 26 met our inclusion criteria. As compliance rates with recommendations were not available in three of them, we
analyzed information from 23 studies (17-39) accounting for 143 recommendations for clinical practice. These 143 "messages" addressed 70 different aspects of medical practice, ranging from the appropriate use of medical or surgical procedures to issues of quality of medical records (a full list of the recommendations is available upon request).

For each recommendation, information on its clinical practice area, year of release, content, and observed compliance rate were gathered. The year in which each identified study had been carried out and its design were also collected. For studies which provided a "before-after" assessment we included only the "after" measurement of compliance. Studies were classified according to their data source: studies using physician self-report data versus those assessing compliance rates with chart audit or administrative data. In addition, we recorded whether the compliance data were collected prospectively or retrospectively.

Given both the potential upward bias of self-report data on compliance, and the fact that the denominator of compliance rates is physicians for self-report but cases for the other data sources, we present the results for these different study designs both in aggregate and separately.

Clinical recommendations were categorized according to the following variables: type of procedure addressed (diagnostic, physical examination, medical treatment, surgical treatment, and other); area of practice (cardiology, oncology, preventive medicine, dental care, obstetrics and gynecology); and characteristics of recommendation.

For characteristics of recommendation, we adapted the five category taxonomy proposed by Rogers (8,11) -- as modified in Table I -- in order to classify the clinical recommendations according to those aspects most likely to influence the diffusion of an innovation, where "innovation" was defined as the procedure addressed by each recommendation. Lack of information in the studies meant that we were unable to assess the first two of Rogers' categories -- "relative advantage" and "compatibility" --
Table I: Definitions adopted in order to classify recommendations according to complexity, trialability and observability of the endorsed procedures.

<table>
<thead>
<tr>
<th>DEFINITION</th>
<th>EXAMPLES of PROCEDURES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Complexity</strong></td>
<td></td>
</tr>
</tbody>
</table>
| High: when a practitioner with usual training and skills, working in an average setting, a) perceives it to be difficult to acquire the skills for or to understand; or b) has no direct control over any resource changes required for implementation. | Radiotherapy treatment for cancer.  
Vaginal delivery for breech presentation. |
| Low: otherwise.  | Antibiotic prophylaxis for bacterial endocarditis.  
Reporting information on the extent of disease on clinical charts. |
| **Trialability** |                                                             |
| High: when a practitioner with usual training and skills, working in an average setting, is able to experiment with it on a limited basis (because either the required skill acquisition or the increase in resource requirements are limited). | Surgical management for breast cancer.  
Pap-test for cervical cancer screening. |
| Low: otherwise.  | Fetal monitoring.  
Specialist consultation for the management of rectal, breast and lung cancer. |
| **Observability**|                                                             |
| High: when a practitioner with usual training and skills, working in an average setting, is able to obtain timely feedback on any impact of the intervention on patients' outcomes. | No cesarean section after previous cesarean.  
Monitoring of toxicity in delivering chemotherapy. |
Rectal, pelvic and breast physical examination. |
which represent an interaction between the "message" and "setting" categories of potential influence. Thus, the recommendations were classified as being high or low on "complexity", "trialability" and "observability" in a typical clinical setting. All the clinical recommendations were classified independently by each of the authors, and the level of agreement was measured with the kappa statistic (40). Overall agreement on the above described classification was 80% (kappa=42%), 87% (kappa=54%) and 82% (kappa=64%) for complexity, trialability and observability respectively. After discussions to solve discrepancies, disagreement persisted only on 7% of recommendations for complexity (kappa=79%) and on 3% for trialability (kappa=91%) and observability (kappa=94%). The 13 recommendations on which no agreement was possible on the basis of the definitions were excluded from this portion of the analysis.

We had a priori expectation that recommendations concerning the use of procedures with high complexity, low trialability and/or low observability would have the lowest compliance rates, and those concerning procedures with low complexity, high trialability and/or high observability the highest.
Statistical Analysis

For the purpose of our analysis, compliance rate was considered as a continuous variable ranging from 0% to 100% with an approximately normal distribution. The validity of this assumption was checked with the Kolmogorov-Smirnov's (K-S) one sample test (41).

As most of the studies accounted for more than one recommendation, a two-way analysis of variance model for unbalanced design (42) was used in order to compare the mean compliance rates according to the different characteristics of the recommendations while accounting for the effect of the study factor. This procedure allowed us to select those characteristics of the recommendations more likely to explain the observed variability in compliance rates.

A multiple linear regression (42) was then used to assess the overall amount of variability in compliance rates which could be accounted for by all the pre-selected characteristics of the message, while controlling for study, source of data (physicians' report vs other) and clinical practice area (cardiology and oncology vs other). We used as weights in the regression model the number of providers (for self-report studies) and the number of cases (for chart audit and administrative data studies) on which each compliance rate was based.
RESULTS

As Table II shows, of the 143 recommendations identified 37% came from surveys providing physicians' self-reported compliance (n = 53), 9% had compliance rates assessed using administrative data (n = 13), and 54% were assessed with chart audit data (n = 77). The elapsed time from release of the guideline to the time at which compliance assessments were undertaken varied from one year or less (36%), through two years (33%), three years (1%), and greater than three years (20%); the longest time was nine years, and the mean across all studies was 2.9 years. Although there was a trend for higher compliance rates with more elapsed time to assessment, this was non-significant (p=0.13).

<table>
<thead>
<tr>
<th>Source of Information</th>
<th>Only after</th>
<th>Before-After Retrospective</th>
<th>Before-After Prospective</th>
</tr>
</thead>
<tbody>
<tr>
<td>No ( %)</td>
<td>No ( %)</td>
<td>No ( %)</td>
<td></td>
</tr>
<tr>
<td>Self-report</td>
<td>41 (29)</td>
<td>- ( -)</td>
<td>12 ( 8)</td>
</tr>
<tr>
<td>Administrative Data</td>
<td>9 ( 6)</td>
<td>1 ( 1)</td>
<td>3 ( 2)</td>
</tr>
<tr>
<td>Clinical Chart Audit</td>
<td>63 (44)</td>
<td>14 (10)</td>
<td>- ( -)</td>
</tr>
</tbody>
</table>
Overall, the mean compliance rate was 54.5% (95%CI 50.2 - 58.9; median 58%; range 0.3%-99%). Figure 1 presents this distribution, regardless of practice area, for the physician and case denominator calculated compliance rates. This proved to be not statistically different from a normal distribution when tested with the K-S test (p=0.30).

Figure 1:  Frequency distribution of compliance rates with the 143 identified official recommendations for clinical practice.
As expected, compliance rates from studies using self-reported information (physician denominators) tended to be higher (mean 59%, median 67%, range: 3-99%; K-S test, p = 0.31) than those from the other two data sources (mean 52%, median 51.5%, range: 0.03-97%; K-S test, p = 0.72). The differences were not, however, statistically significant (Figure 2).

According to the area of clinical practice (Figure 2), the mean compliance rates ranged from 43.3% to 63.6%. Recommendations on cardiovascular and cancer patients' management had significantly higher compliance rates (63.6% and 62.5%, respectively) than those for preventive care (46.2%), dental care (45.8%) or obstetrics and gynaecology (43.3%) (p < 0.001). The mean elapsed time from the release of recommendations to compliance assessment did not explain any of these differences and was not significantly different according to the area of clinical practice.

Figure 2

![Figure 2: Mean compliance rates (with 95% confidence intervals (CI) with the 143 identified recommendations, according to practice area and data source.](image-url)
No major differences in compliance rates emerged according to the type of procedure for each recommendation (Figure 3). The source of data used in the studies (self-reported vs other) influenced unevenly the observed rate, being more evident for recommendations concerning physical examination (75.7% vs 47%) and the use of appropriate surgical treatments (62% vs 52%) (Figure 3).

Figure 3: Mean compliance rates (with 95% confidence intervals (CI) with the identified recommendations, according to type of procedure and data source.

* includes recommendations concerning tests and procedures for diagnostic purposes or for cancer staging.

** includes recommendations concerning biopsies and all other surgical procedures for diagnostic or staging purposes (i.e. second look laparotomy for ovarian cancer).

*** includes recommendations concerning the quality of information reported on clinical charts and those supporting discussion with patients of alternative courses of action.
The characteristics of the recommendations did appear to have some predictive capacity for compliance rates (Figure 4). Recommendations judged to be of high complexity had significantly lower compliance rates (41.9%) than those judged to be of low complexity (55.9%) \( (p=0.05) \). Recommendations judged to be high on trialability showed a significantly higher compliance rate (55.6%) than those judged as low on trialability (36.8%) \( (p=0.03) \). High versus low observability of the recommendation did not, however, demonstrate any significant difference in compliance rate (54.6% versus 52.4%). When analyzed separately for the source of the data (self-report versus other), the same pattern was evident except for observability, where only the self-report data showed a slight trend in the predicted direction of higher compliance for high observability.

Figure 4

<table>
<thead>
<tr>
<th>Complexity</th>
<th>Number of Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>108</td>
</tr>
<tr>
<td>High</td>
<td>24</td>
</tr>
<tr>
<td>Low Trialability</td>
<td>18</td>
</tr>
<tr>
<td>High Trialability</td>
<td>114</td>
</tr>
<tr>
<td>Low Observability</td>
<td>78</td>
</tr>
<tr>
<td>High Observability</td>
<td>52</td>
</tr>
</tbody>
</table>

Figure 4: Mean compliance rates (with 95% confidence intervals (CI), with clinical recommendations classified according to complexity, trialability and observability.
To evaluate the potential contribution of all the analyzed characteristics of the message accounting for type of data source (self-report vs other) and study, we did a multivariate analysis (Table III). Overall the characteristics of the message listed in Table III explained 23% of the variance. The major contributing variables with significant coefficients were: complexity (coeff.: -0.24), and clinical practice area (coeff.: 0.12). When analyzed separately the non-self-report, or case-denominator compliance rates, yielded a much higher R-squared (47.5% of variance explained), but the pattern of variables with significant coefficients was left unchanged.

Table III: Results of the multiple regression analysis assessing the strength of the association between compliance rate and some key characteristics of the message of the recommendations.

<table>
<thead>
<tr>
<th>Variables</th>
<th>All the studies</th>
<th>Excluding self-reported studies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(130)</td>
<td>(82)</td>
</tr>
<tr>
<td>STUDY</td>
<td>0.008</td>
<td>0.004</td>
</tr>
<tr>
<td>DATA SOURCE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=Self-report</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0=Chart audit/Administrative</td>
<td>0.12*</td>
<td>not included</td>
</tr>
<tr>
<td>AREA OF PRACTICE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=Cardiology/Oncology</td>
<td>0.01*</td>
<td>0.13 *</td>
</tr>
<tr>
<td>0=Otherwise</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMPLEXITY</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=High</td>
<td>-0.24**</td>
<td>-0.32 **</td>
</tr>
<tr>
<td>0=Low</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRIALABILITY</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=High</td>
<td>0.03</td>
<td>0.14</td>
</tr>
<tr>
<td>0=Low</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.23</td>
<td>0.475</td>
</tr>
</tbody>
</table>

( )= number of recommendations
* = p < 0.05
** = p < 0.01
DISCUSSION

We found a high degree of variation in compliance with practice guideline recommendations, and an average compliance rate of just over 50%. This study, however, did not evaluate the quality of the recommendations in relation to the existing research evidence. Thus some of the lack of compliance might represent appropriate deviations that better reflect good quality care than practices which followed the recommendations. It is important to note, however, the official nature of the guidelines, their objective of synthesizing evidence to reflect the best quality of care, and the fact that there was a lengthy mean time available in the studies for adjustments of practice to be made to comply with the message (nearly three years). These considerations suggest that the observed compliance rates are indicative of much room for improvement in the quality of currently delivered medical care.

The variability did allow us to make some assessment of the potential predictors of compliance rates. We focused on the characteristics of the messages represented by the 143 recommendations, particularly each message's relationship to those variables isolated as important by the diffusion of innovations literature (8). We wished to see how empirically useful this literature might be in providing prediction of the compliance rate with a recommendation.

Using only the characteristics of the message, i.e. ignoring the potential influence of the setting in which compliance was assessed or the mediating effect of the particular provider characteristics and predilections, we were able to explain nearly 50% of the variation in chart-review based compliance rates. Messages (recommendations) concerning high complexity and low trialability obstetric or preventive interventions are likely to result in low levels of compliance. This is in contrast to recommendations for cardiovascular or oncologic interventions of low complexity and high trialability, which are likely to result in much higher levels of compliance.
Perhaps surprisingly we found no influence of observability on compliance rates. We say surprising because the definition of this diffusion factor (see Table I) relates it potentially to clinical relevance. The absence of clinical relevance in practice guidelines has often been cited as a major explanation for observed lack of compliance (43). Insofar as clinical relevance is represented by our definition of observability, this study did not substantiate this hypothesized relationship. We would note, however, that our inability to assess relative advantage of a procedure resulted in omission of the variable likely to be most closely related to perceived clinical relevance. We also found no relationship of the type of procedure addressed in the message to the subsequent compliance rate.

Although we have no definitive explanation for why the cardiology and oncology areas of practice show significantly higher compliance rates, we offer the following hypothesis. Practitioners from both these specialty areas are more likely than the average physician to be located in a secondary or tertiary care setting in which there is some ongoing transfer of research knowledge. This is in contrast to the more isolated ambulatory or primary care settings in which family physicians, dentists, and obstetrician-gynaecologists are likely to practice. Thus the cardiology and oncology recommendations may well be received in an environment or culture that is more conducive to consideration of the relevant and synthesized research information represented by practice guideline recommendations.

Although about half of the variance in compliance rates could be explained on the basis of diffusion and other characteristics of the message, half was not explained. This is not surprising given our omission of consideration of any of the other categories of potentially influential factors listed by Winkler et al. (7) -- the source of the message, the channel of communication, the clinical setting, and the nature of the receiving audience. To a large extent the source and the channel were held constant across the studies, and would not therefore be expected to contribute to an explanation of variance in compliance. The source was preselected to be an official organization from more
than a local environment. The channel was publication and dissemination through the research and educational medical literature.

Both the diffusion of innovations literature and the growing physician behaviour change literature would, nevertheless, support the idea that the setting and the audience are likely to be quite influential on compliance rates. This study does not enable us to differentiate between the relative contributions to higher compliance rates for cardiology of the characteristics of the cardiologic procedures (the message), the cardiologists (the receiving audience), where cardiologists work (the setting), and the interactions among all three. The fact that the judgement of complexity for the procedure involved in a recommendation was the single most effective predictor of compliance suggests that our definition for this characteristic of the message may have taken account of some of the potential interactions with setting and audience. In any event, more accurate assessments of prediction factors for compliance with guideline recommendations must clearly involve consideration of these setting and audience factors. These will necessarily involve prospective studies that can specify setting and audience characteristics with an accuracy that was beyond the capability of this passive review of existing reports.

Nevertheless, even this limited study does provide some implications for the burgeoning use of practice guidelines. For instance, and assuming that improved compliance rates translate directly into improved patient out-comes, one objective might be to ensure that the quality of care seen in cardiology and oncology is at least the minimum level for all specialties. In this case limited resources for the active implementation of practice guidelines may best be dedicated to preventive services and obstetrics-gynecology rather than to cardiology and oncology. Alternatively, if the objective is to "pick the winners" for use of limited implementation resources, then efforts may best be directed at further improving quality in cardiology and oncology.

Also, the nature of efforts may be different for high complexity/low trialability messages compared to low complexity/high trialability messages. For the former much
greater effort might go into predisposing practitioners to even consider making changes in their delivery of care before embarking on more local efforts at encouraging and enabling the changes to occur. For the latter perhaps less effort is needed to predispose practitioners towards change, under the belief that practitioners are already receptive to this kind of alteration in their practices, and efforts can more quickly move to a focus on strategies such as opinion leaders to enable change at the local level (44).

Without further research, however, it is still not clear whether, even with a highly "implementable" message, efforts to enable change according to a practice guideline are best directed at the receiving audience of practitioners, the incentives and structure of the settings in which they work, or the interaction between the two. These types of questions presumably will be the focus of some of the numerous research projects that are now evaluating various dissemination strategies under the Agency for Health Care Policy and Research's medical effectiveness and practice guideline research programs. The results of this study provide some assistance to these investigators by better defining the relationship between characteristics of the practice guideline's message and expected compliance rates.
REFERENCES


