Clinical trial: evaluation of a clinical decision-support model for upper abdominal complaints in primary-care practice

N. HOROWITZ*, †, M. MOSHKOWITZ*, ‡, M. LESHNO‡, §, J. RIBAK*, S. BIRKENFELD**, G. KENET†† & Z. HALPERN*, ‡

*Department of Gastroenterology, Tel-Aviv Sourasky Medical Center, Tel-Aviv; †Department of Epidemiology, Tel-Aviv University, Tel-Aviv; ‡Faculty of Management, Tel-Aviv University, Tel-Aviv; §Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv; ††School of Public Health, Tel-Aviv University, Tel-Aviv, Israel; **Briuth Clalit Health Service, Israel; ††Maccabi Health Service, Israel

Correspondence to:
Dr N. Horowitz, Department of Gastroenterology, Tel-Aviv Sourasky Medical Center, 6 Weizmann St, Tel Aviv 64239, Israel.
E-mail: noyah@tasmc.health.gov.il

Publication data
Submitted 6 June 2007
First decision 2 July 2007
Resubmitted 12 August 2007
Accepted 21 August 2007

SUMMARY

Background
Gastro-oesophageal reflux disease (GERD) and dyspepsia affect 25–40% of the general population. In the absence of alarm symptoms, the current recommended policy in young dyspeptic patients is a ‘test and treat’ strategy for Helicobacter pylori; in GERD patients, a therapeutic trial with proton pump inhibitors is the treatment of choice.

Aim
To create a short and simple clinical algorithm, for the diagnosis and treatment of patients with upper gastrointestinal complaints.

Methods
The clinical usefulness and cost-effectiveness of the new algorithm were evaluated in a controlled clinical trial, held in primary-care clinics in Israel. Clinical and economical treatment outcomes were evaluated after 1, 3 and 6 months comparing doctors who used the algorithm (cases) vs. those who did not (controls).

Results
78 cases and 54 controls completed the 6 months of follow up. The improvement in symptom severity and quality of life was greater in the cases than in the controls (P < 0.05). General practitioner clinics visits (P = 0.04), gastroenterology clinics visits (P = 0.02) and medication costs (P = 0.004) were all significantly reduced among cases. Controls underwent also more imaging tests (computerized tomography, ultrasound and X-ray) and endoscopies. The average cost for 6 months’ treatment and follow-up was $US 199 for cases compared with an average of $US 336 in the control group.

Conclusion
The use of a clinical decision-support tool can facilitate and promote the implementation of management guidelines by general practitioners. The short algorithm presented in the study was found to be useful and easy to apply in clinical practice. Its effectiveness can be further increased by implementing it in computerized medical systems.
INTRODUCTION

Upper abdominal complaints are prevalent, affecting 25–40% of the general population.\(^1\)\(^-\)\(^3\) The assessment and treatment of these symptoms place a substantial burden on patients and the healthcare delivery system, especially with the introduction of new drugs and the increased use of diagnostic interventions.\(^4\)\(^,\)\(^5\)

Healthcare suppliers and scientific committees use evidence-based methods and consensus expert panels, to create clinical guidelines to reduce healthcare costs and improve treatment outcomes in dyspeptic patients. Many algorithms and treatment protocols were, indeed, published, but the inherent difficulty is in getting doctors to implement them. The guidelines are often voluminous and complicated, and sit on a shelf gathering dust.\(^6\) We believe, however, that a primary stratification of patients into main diagnostic and treatment groups, based on their symptoms, might be less costly and easier to apply in primary-care settings.

In the absence of alarm symptoms, the most important and the least expensive mechanism for diagnosing gastro-oesophageal reflux disease (GERD) is the typical manifestations, such as predominant heartburn or acid regurgitation, completed by a therapeutic trial with proton pump inhibitors (PPI).\(^7\)\(^-\)\(^14\) On the other hand, the current recommended approach in young dyspeptic patients, presenting with a less clear clinical profile, is a ‘test and treat’ strategy for Helicobacter pylori.\(^11\)\(^,\)\(^12\)\(^,\)\(^15\)\(^-\)\(^18\)

To help and guide primary-care doctors in choosing the appropriate diagnostic work-up for patients presenting with upper abdominal complaints, data mining techniques have been used to develop a simple algorithm.\(^19\) The algorithm was designed to classify patients into GERD and non-GERD, based on six common dyspeptic symptoms.

The aim of this study was to evaluate the clinical and economic effectiveness of the algorithm in primary-care clinics.

METHODS

The study was carried out on patients presenting with upper abdominal complaints, at 16 primary-care clinics randomly divided into two groups: a treatment group (cases) and a control group (controls). At the eight clinics of the treatment group, patients were diagnosed and treated according to the algorithm, and at the eight clinics of the control group, diagnosis and treatment were at the doctor’s discretion, with no guidance. The clinics were selected from a list of facilities located in various socioeconomic areas: three of the eight in each group were located in well-established neighbourhoods, three in urban (down-town) areas and two in low socioeconomic areas. The clinical and economic outcomes of the two groups were assessed and compared.

Study design

The management algorithm consisted of a 6-symptom diagnostic questionnaire, followed by a non-invasive diagnostic test. The development and validation of the questionnaire is described elsewhere\(^19\) (Appendix 1). Patients classified by the questionnaire as Reflux patients – received a therapeutic trial with PPIs, while the others were managed by a ‘test and treat’ strategy for H. pylori.\(^15\) The management chart is illustrated in Figure 1.

Yes       No

Alarm symptom is present?

Gastroscopy

Diagnostic questionnaire

Reflux disease

PPI treatment

Non-reflux

\[^{13}\]C Urea Breath test for Helicobacter pylori (test and treat strategy)

* Negative patients will be managed empirically according to their symptoms, or referred to subspecialty consultation.

Figure 1. Management algorithm for cases.
Subject selection

Inclusion criteria

To be eligible for study entry, patients were required to be at least 18 years of age, presenting with an uninvestigated upper abdominal symptom, such as heartburn, epigastic pain or other discomfort located in the upper gastrointestinal tract.

Exclusion criteria

Patients who underwent upper gastrointestinal endoscopy during the preceding year could not participate in the study.

Efficacy and cost assessments

Outcome measures were assessed at baseline and on months 1, 3 and 6 from baseline.

The primary efficacy end-point was relief (or resolution) of upper abdominal symptoms, graded by the subjects, using a 10-point scale for global assessment of symptom severity. Another outcome measure was the reduction in symptom interference with daily activities, assessed globally by the subject using a 5-point Likert scale.

All healthcare resources consumed by patients during the 6 months following initial evaluation were recorded and analysed as well. These include the number of consultations with primary-care doctors, referrals to specialists, medication use and diagnostic tests performed by each patient, for upper abdominal symptoms. Drug prices were based on Israeli Ministry of Health (MOH) price list for pharmaceutical products, updated to September 2004. The costs associated with primary-care consultations, referrals and investigations were based on the MOH price list for open-access healthcare services, updated to September 2004 as well.

Work loss – because of upper abdominal symptoms – was also evaluated by the subjects at each contact (months 1, 3 and 6 from baseline) regarding the preceding period.

The study was conducted in accordance with the declaration of Helsinki and was consistent with Good Clinical Practice. Ethics committee approval was obtained before the study commenced and all patients signed informed consent.

Statistical analysis

We estimated the sample size for the following outcomes: symptoms severity score (on a 0–10 points Likert scale), quality of life (on a 0–5 points Likert scale) and total costs. Assuming a mean decrease of 5 ± 3 points in symptoms severity score in the treatment group compared to the control group, 47 subjects were required in each group to ensure adequate power 90%, with a two-sided alpha of 0.05. Assuming a mean decrease of 2 ± 1 points in the quality of life scale, 20 subjects were required in each group, and assuming a mean reduction of $US 150 in the total costs, 20 subjects were required in each group, to ensure the same power and alpha. Using adjustment for multiple comparisons and a drop-out rate of 10%, 52 subjects were required in each group.

All variables were analysed using the intention-to-treat populations, which included all recruited patients who agreed to follow-up. Statistical tests were conducted using SPSS software, version 12.0. All P-values were two-sided, and considered significant at 0.05.

RESULTS

One hundred and thirty-eight consecutive patients were enrolled into the study – 80 subjects in the treatment group and 58 subjects in the control group. Two of the cases did not complete the follow-up and therefore were not included in the data analysis (one of them withdrew his consent and the other was lost to follow-up). Four controls were excluded as well (two of them were not eligible to begin with, as they could not speak Hebrew, another subject withdrew his consent, and a forth patient was lost to follow-up).

All clinicians were monitored to determine weather they were using the algorithm correctly. Although one case of the treatment group was not treated according to the guidelines, it was not omitted from the data analysis (one of them withdrew his consent and the other was lost to follow-up). Four controls were excluded as well (two of them were not eligible to begin with, as they could not speak Hebrew, another subject withdrew his consent, and a forth patient was lost to follow-up).

All clinicians were monitored to determine weather they were using the algorithm correctly. Although one case of the treatment group was not treated according to the guidelines, it was not omitted from the data analysis (one of them withdrew his consent and the other was lost to follow-up). Four controls were excluded as well (two of them were not eligible to begin with, as they could not speak Hebrew, another subject withdrew his consent, and a forth patient was lost to follow-up).

Patients who were managed according to the algorithm experienced a significantly greater relief in symptoms severity than the controls (Table 2). The relief in symptoms was also faster among cases. One month after baseline, their symptoms improved by 3.5
points (out of 10), in comparison with 1.98 points among controls (Figure 2). And, the most important: while symptom severity among cases continued to decrease even after 6 months of follow-up, among controls, the improvement in symptoms almost stopped after 3 months, leaving the patients still presenting with symptoms, but apparently without an effective treatment.

The change in symptom severity as a function of time in both groups was found to be significant (Greenhouse-Geisser test, \( P < 10^{-5} \)), and the difference between cases and controls was highly significant as well (\( P < 0.0001 \)).

Improvement in daily activities was greater in cases than in controls (2.2 vs. 1.7 points out of five, respectively). Even though the 5-point scale is less sensitive to change, the difference between cases and controls is almost significant (\( P = 0.09 \), Mann–Whitney \( U \)-test; Figure 3).

A significant difference was also found in management costs between the two groups of patients (Table 2). The average cost of 6 months’ treatment and follow-up was $US 199 in cases compared with an average of $US 336 in controls. Even after omitting all patients with alarm symptoms (in whom treatment might be more expensive), the difference in management cost between cases and controls remained clear and significant (a cost of $US 198 ± 24 in cases and $US 315 ± 49 in controls, \( P = 0.006 \)). The reduced treatment-cost among cases – compared to controls – represents different health resources savings, especially in medication costs (almost half among cases), specialist referrals (almost half among cases, as well) and primary-care consultations (Table 3).

As expected, more \( ^{13} \text{C} \)-urea breath tests (UBT) were performed among cases (as all dyspeptic patients with

---

**Table 1. Baseline age and gender of the study population**

<table>
<thead>
<tr>
<th>Age &amp; gender</th>
<th>Treatment group</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age (±s.d.)</td>
<td>52.1 (16.4)</td>
<td>53.0 (19.1)</td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
<td>25 (44%)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>45 (56%)</td>
</tr>
</tbody>
</table>

Significance (two-tailed; *P*-value)

* Independent \( t \)-test; † \( \chi^2 \)-test.

---

**Table 2. Outcome measures improvement at 6-month follow-up**

<table>
<thead>
<tr>
<th>Outcome measure</th>
<th>Treatment group ((N = 78; \text{mean ± s.d.}))</th>
<th>Control cases ((N = 54; \text{mean ± s.d.}))</th>
<th><em>P</em>-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improvement of symptoms severity score after 6 months (0–10 point scale)</td>
<td>5.5 ± 2.9</td>
<td>3.0 ± 2.9</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Improvement interference of symptoms with daily activities (0–5 point scale)</td>
<td>2.2 ± 1.1</td>
<td>1.6 ± 1.3</td>
<td>0.009</td>
</tr>
<tr>
<td>Cumulative total cost ($US)</td>
<td>199 ± 211</td>
<td>337 ± 300</td>
<td>0.002</td>
</tr>
</tbody>
</table>

* Mann–Whitney \( U \)-test or \( t \)-test.

---

Figure 2. Symptom severity during 6-month follow-up.
no alarm symptoms or typical reflux symptoms were
directed to $^{13}$C-UBT). The difference between the two
groups was, however, trivial and not statistically sig-
nificant (Table 3).

About 24% of the cases and 37% of the controls
underwent gastroscopy. In the absence of alarm symp-
toms, theoretically unnecessary endoscopies were per-
formed by 32% of the controls, compared to 19% of the
cases (a clear difference, although not significant).

Thirty-four employed controls lost a total of 32
workdays, because of their upper abdominal symp-
toms, compared with only six workdays lost among
51 employed cases. The average workday loss was
0.94 ($\pm 2.6$) days per person in the control group com-
pared with a 0.12 ($\pm 0.43$) day per person among the
cases ($P = 0.07$, t-test).

### DISCUSSION

Previous studies have evaluated the test and treat
strategy in dyspeptic patients$^{23, 24}$ as well as PPI ther-
apeutic trial in GERD patients, separately.$^{25–27}$ However,
patients usually present with a mixture of symptoms,
placing difficulties in classifying them into the appro-
priate diagnostic and therapeutic category.

The present prospective study was designed to
address this issue, and evaluated the clinical and eco-
nomical effectiveness of a symptom-based algorithm
for the management of all types of uninvestigated dys-
pepsia in primary-care set up. We found that general
practitioners (GP) who used this algorithm improved
their performance in symptom classification. As a
result, clinical outcomes were improved, and a sub-
stantial reduction in healthcare costs and medication
use was achieved.

An important finding in the present study is a sig-
nificant reduction in medication use among cases. This
is worth noting even more considering the fact that
cases with reflux-type symptoms were treated initially
with PPIs as the treatment of choice. Most probably,
the increased medications use among controls is a
result of inappropriate use of H. pylori testing and
eradication. Chey et al.$^{28}$ have recently reported a gen-
eral confusion among primary-care doctors about the
relationship between GERD and H. pylori infection,
and that 80% of them tested for the infection in at
least some patients who had only symptoms of GERD.
Niv et al.$^{29}$ have also reported a significant gap in
knowledge between gastroenterologists and family
doctors regarding the diagnosis and treatment of
GERD.

### Table 3. Cumulative costs and medical resources consumption among study population groups

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Treatment group $(N = 78; \text{mean} \pm \text{s.d.})$</th>
<th>Control cases $(N = 54; \text{mean} \pm \text{s.d.})$</th>
<th>$P$-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of GP visits</td>
<td>1.7 ± 0.8</td>
<td>2.0 ± 0.9</td>
<td>0.04</td>
</tr>
<tr>
<td>Number of specialist referrals</td>
<td>0.38 ± 0.6</td>
<td>0.74 ± 0.9</td>
<td>0.02</td>
</tr>
<tr>
<td>Cost of medications ($$$)</td>
<td>78 ± 343</td>
<td>146 ± 699</td>
<td>0.004</td>
</tr>
<tr>
<td>Number of imaging diagnostic tests†</td>
<td>0.09 ± 0.33</td>
<td>0.19 ± 0.44</td>
<td>0.18</td>
</tr>
<tr>
<td>Number of $^{13}$C-urea breath tests‡</td>
<td>0.38 ± 0.48</td>
<td>0.3 ± 0.46</td>
<td>0.3</td>
</tr>
<tr>
<td>Number of gastroscopies‡ for patients aged ≤45 years</td>
<td>0.24 ± 0.43</td>
<td>0.35 ± 0.48</td>
<td>0.24</td>
</tr>
<tr>
<td>Number of gastroscopies‡</td>
<td>0.04 ± 0.2</td>
<td>0.35 ± 0.48</td>
<td>0.01</td>
</tr>
</tbody>
</table>

* Independent $t$-test; † Computerized tomography, abdominal ultrasound, abdominal X-ray, barium meal; ‡ Fisher exact test.
Most doctors can barely keep pace with the rapid advances in healthcare knowledge and clinical guidelines, which in many cases are complicated and difficult to implement. Assimilation and implementation of new strategies are possible only when a well-designed intervention is used, such as clinical decision-support tools. The management algorithm presented here is based on a short, structured interview to collect patient symptoms and to classify patients to GERD vs. non-GERD evidence-based work up. We have found this algorithm simple to administer, easy to implement and significantly effective in primary-care practice.

The results of the present study emphasize an important point which is the urgent need for a clinical education for facilitating the uptake of dyspepsia management guidelines and principals in primary care. Assessing dyspeptic symptoms is a challenge for both gastroenterologists and primary-care doctors. The symptom-driven nature of diagnosis and patient management places particular importance on a good estimation of symptoms including severity, duration and frequency. For this purpose, clinical practice guidelines as well as several dyspepsia and GERD questionnaires have been developed and published in the recent years. However, it has been shown that even well-constructed guidelines have little effect unless supported by an appropriate educational programme and implementation strategies. Banait et al. have shown that educational outreach was more effective than passive guidelines dissemination in changing clinical behaviour in the management of dyspepsia. However, we think that the simple algorithm presented in our study shows a good practicability in the primary-care setting, and this makes it suitable for use in clinical trials and as a guide to patient management in a broader primary-care population.

One of the major criteria used in the algorithm presented here for predicting an underlying malignancy is the existence of alarm features. However, the value of this factor is uncertain. In a recent meta-analysis published recently, Vakil and colleagues found that alarm features have limited predictive value for an underlying malignancy in patients with dyspepsia.

Awareness of the limitation of alarm features in predicting malignancy is needed according to these results. However, currently this is the most useful and applicable strategy that can be used in the management of dyspepsia.

Another potential advantage of this algorithm is that its effectiveness can be further increased by implementing it in computerized medical systems, where it is likely to facilitate active assimilation of guidelines, and decrease the amount of doctor time necessary for appropriate diagnosis and management.

In conclusion, we found that the use of a clinical decision-support tool can facilitate and promote the implementation of management guidelines by GPs. The short algorithm presented in the study was found to be useful and easy to apply in clinical practice.

ACKNOWLEDGEMENT
Declaration of personal and funding interests: None.

REFERENCES
10 Carlesson R, Dent J, Bolling-Sternevald E, et al. The usefulness of a structured questionnaire in the assessment of


APPENDIX: DIAGNOSTIC SCORE

The classification is based on a ‘GERD score’ yielded by summing the severity response of each symptom multiplied by its relative weight. If the score is 30 points or more than the patient will be classified as having GERD. Patients can be also classified as having GERD according to the severity of reflux typical symptoms (heartburn, acid regurgitation or nocturnal reflux). If at least two reflux symptoms are marked as ‘4’ (and up), or at least one reflux symptom is marked as ‘5’ than the patient will be classified to the ‘reflux’ management.

All other patients will be classified to the ‘non-reflux’ work-up.

<table>
<thead>
<tr>
<th>Symptom severity</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>GERD score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heartburn</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>$X_8 =$</td>
</tr>
<tr>
<td>Acid regurgitation</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>$X_5 =$</td>
</tr>
<tr>
<td>Sour oral taste</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>$X(-5) =$</td>
</tr>
<tr>
<td>Aggravation of symptoms after heavy meals</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>$X(-1) =$</td>
</tr>
<tr>
<td>Relief of symptoms by antacids</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>$X_3 =$</td>
</tr>
<tr>
<td>Nocturnal reflux</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>$-$</td>
</tr>
<tr>
<td>Total score</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>$-$</td>
</tr>
</tbody>
</table>

© 2007 The Authors, Aliment Pharmacol Ther 26, 1277–1283
Journal compilation © 2007 Blackwell Publishing Ltd