Research Methods

Development and validation of a new instrument measuring guideline adherence in clinical practice

Marlous F Kortekaas a, *, Marie-Louise E L Bartelink a, Geert J M G van der Heijden b, Arno W Hoes a and Niek J de Wit a

a Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands and b Department Social Dentistry, Academic Center for Dentistry Amsterdam, University of Amsterdam and VU University, Amsterdam, The Netherlands

*Correspondence to Marlous F Kortekaas, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Str 6, 131, Postbus 85500, 3508 GA Utrecht, The Netherlands; E-mail: m.f.kortekaas@umcutrecht.nl

Abstract

Background. Education in evidence-based medicine (EBM) is an important part of the postgraduate training of GPs. Evaluation of its effect on EBM behaviour in daily clinical practice is difficult and instruments are scarce. Working in accordance with guidelines is considered as one of the key indicators of EBM behaviour.

Objective. To develop and validate an instrument assessing guideline adherence of GP trainees in clinical practice.

Methods. We developed an instrument that assesses guideline adherence, taking conscious deviation into account. The instrument assesses guideline adherence on 59 different management decisions (diagnosis \( N = 17 \), therapy \( N = 20 \), referral \( N = 22 \)) for 23 conditions as described in 27 different clinical practice guidelines. We validated this instrument using performance data as collected by third-year GP trainees on three important properties: validity, reliability and feasibility.

Results. Performance data were collected by 76 GP trainees on 12,618 patient consultations with 125,877 different reasons for encounter. Overall, guideline adherence was 82% (95% confidence interval 77–88%). The significant correlation with the national GP knowledge test (\( r = 0.33, P < 0.004 \)) showed the instrument to be a valid instrument. Interrater reliabilities (intraclass correlation coefficient) varied between moderate and excellent (0.64–1.00, \( P < 0.001 \)). The instrument proved feasible with coverage of 24% (\( N = 3082 \)) of reasons for encounter presented to GP trainees and a mean and median time of 1 minute to score a patient consultation.

Conclusion. This instrument proved valid, reliable and feasible to assess guideline adherence among trainees in the clinical primary care setting.

Key words: Evidence-based medicine, general practice, guideline adherence, reliability, validity, vocational education.

Introduction

To date, evidence-based medicine (EBM) is well-integrated in medical curricula and in postgraduate training programmes (1). EBM enables health care professionals to make rational medical decisions by integrating the best available evidence with clinical expertise and patient values (2). EBM is considered one of the core competencies in the postgraduate GP training (3) and is integrated in this training in five steps following the Sicily Statement recommendations: ask, acquire, appraise, apply and assess (4). The GP specialty training in the Netherlands takes 3 years and combines 4 days training in

© The Author 2016. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.
clinical practice with 1 day theoretical education at the GP training institute at the University. The main goal is to teach GP trainees to use EBM into their daily practice. Therefore, it is important to be able to evaluate effects of EBM training in clinical practice (4). There is, however, no agreement on the optimal way to assess EBM behaviour (5). It is suggested that it is best assessed by monitoring clinical management decisions (6), but instruments for this are scarce (5,7).

The clinical content of the postgraduate GP training in the Netherlands is based on the clinical practice guidelines (CPGs) for GPs from the Dutch College of General Practice (8). Presently, there are >100 evidence-based CPGs that are the result of a structured process based on collection and assessment of research evidence, and expert opinion (8). Working in accordance with these guidelines is considered as one of the key indicators of practising EBM, and monitoring guideline adherence is a way to assess EBM behaviour during professional training (9). We report on the development of an instrument to assess guideline adherence and its validation among a cohort of GP trainees at the UMC Utrecht.

Methods

Instrument

Description of the instrument

This instrument assesses rational guideline adherence of GP trainees using patient consultation data. Rational adherence means clinical management in accordance with key recommendations in professional guidelines, thereby allowing motivated deviation in individual patients. The instrument assesses guideline adherence on 59 different management decisions (on diagnosis N = 17, on therapy N = 20 and on referral N = 22) for 23 prevalent conditions, as described in 27 different CPGs.

Development process

We started with selecting quality indicators from an existing quality assessment instrument in primary care that is used to monitor guideline adherence among GPs (10,11). We selected quality indicators that measure whether medical decisions adhered to the guideline recommendations. For pragmatic reasons, we focused on those medical decisions that do not require extensive knowledge of medical history or prior medical decisions (12). To increase feasibility, we selected quality indicators for disorders frequently presented to GP trainees (13). This resulted in 19 quality indicators for 17 disorders (24% of all quality indicators in the existing quality assessment instrument).

Based on the outcome of expert sessions (discussions with the academic GPs that were involved in the development of the original guideline), we adjusted the original quality indicators to be able to assess management decisions in individual patient consultations. An example of such an adjustment is shown in Table 1. Pilot testing on anonymized patient data collected by GP trainees demonstrated that the 19 adjusted quality indicators assessed decisions in clinical management of ~20% of all patient consultations. Because we considered 20% low, we decided to extend the number of indicators in the instrument. Since in the first phase each quality indicator addressed only one management decision (on diagnosis, therapy or referral), we extended the instrument with assessment of additional management decisions from the same guidelines. Moreover,

Table 1. An example of an adjustment of an existing quality indicator (11) to an assessment item as included in the developed instrument used to assess guideline adherence

<table>
<thead>
<tr>
<th>Original quality indicator</th>
<th>Numerator: Children &gt;2 years old with AOM who did not receive AB. Denominator: All children &gt;2 years with AOM.</th>
<th>Percentage of children &gt; 2 years with AOM that did not receive AB.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOM in children &gt;2 years (14)</td>
<td>↓</td>
<td>-1 AB</td>
</tr>
<tr>
<td>Adjusted quality indicator (assessment item)</td>
<td>AOM in children (14) No AB, unless indicated. When AB indicated, according to the schedule as provided in the GP guideline (15).</td>
<td>AB prescribed when not indicated or not according to schedule in GP guideline, and no reason mentioned (e.g. parent’s wish, allergy). Quinolones and/or lidocaine prescribed as eardrops.</td>
</tr>
<tr>
<td>N = 22</td>
<td>No AB</td>
<td>No AB prescribed when indicated.</td>
</tr>
<tr>
<td>?</td>
<td>No AB prescribed despite otorrhea at first presentation, and no advice given to come back when otorrhea lasts for &gt;1 week.</td>
<td></td>
</tr>
<tr>
<td>0 AB</td>
<td>AB prescribed as eardrops (not being quinolones or lidocaine).</td>
<td></td>
</tr>
<tr>
<td>1 AB</td>
<td>AB prescribed when indicated, and according to schedule in GP guideline, or deviated from guideline with reason (e.g. parent’s wish, allergy).</td>
<td></td>
</tr>
<tr>
<td>No AB</td>
<td>No AB prescribed when not indicated.</td>
<td></td>
</tr>
</tbody>
</table>

AB, antibiotics; AOM, acute otitis media.

1AB indicated according to GP guideline: (i) >3 days of complaints with fever; (ii) <3 days of complaints, but very ill or quickly becoming more ill; (iii) patient in risk group, as defined in GP guideline; (iv) otorrhea at first presentation or for >1 week; (v) AOM in both ears and child <2 years old.

2AB schedule according to GP guideline: (i) Amoxicillin (7 days); (ii) Azithromycin (3 days) or co-trimoxazole (5–7 days) in case of contra-indications for amoxicillin; (iii) Augmentin in case of otorrhea in ill children with ear tubes.
we expanded the number of guidelines involved and designed new assessment items on management decisions for other disorders that were frequently presented to GP trainees. The final set of assessment items on management decisions was again extensively discussed with academic GP experts in one-to-one sessions.

Scoring guideline adherence
Guideline adherence was scored by two independent researchers (Yvonne Spoormans and Lisanne Welink) for every decision in clinical management (diagnosis, therapy and referral) on a 3-point scale: (−1) not in accordance with the guideline and no reason mentioned to deviate from it; (0) debatable if in accordance with the guideline due to insufficient or contradicting information; and (+1) in accordance with the guideline or not in accordance but with rational motivation, for example, in accordance to patient’s wish or clinical expertise. In case of disagreements, a third researcher (Jeroen van Duijn) made a final decision.

Validation process
Data collection
We validated the instrument in data from the PINET study (Personalized INtegrated Evidence-based medicine teaching for Trainees in general practice), a prospective, cluster randomized controlled trial, in which third-year GP trainees from Utrecht University were allocated to either an integrated EBM training programme or a stand-alone EBM training programme. Demographics and baseline characteristics of all participants were collected. We used data from two consecutive measurements, at the beginning and end of the third year of the postgraduate GP training programme. GP trainees collected the following data from the electronic medical records (EMRs) in the GP trainee practice in logs: gender, age, reason for encounter, medical history, results of physical examination, diagnosis [International Classification of Primary Care (ICPC) code (16)] and treatment or referral. If a patient presented multiple reasons for encounter during one consultation, multiple ICPC codes were allocated. We used ICPC codes to select patient consultations regarding disorders described in the instrument (Table 2). We excluded the following consultations from assessment: (i) a follow-up consultation, as prior management decisions affect the GP trainee’s decision; (ii) a patient consultation that fits in multiple guidelines addressing a single reason for encounter (e.g. common cold is addressed by the guidelines on ‘acute cough’, ‘rhinosinusitis’ and ‘acute sore throat’); (iii) missing data on age or gender, when relevant (e.g. in a patient consultation about asthma where the management decision depends on the age of the patient) (17,18); and (iv) a patient consultation in which the aim of the consultation does not match the selected assessment items (e.g. a patient with diabetes mellitus who wants to talk about the social implications of the disease).

Psychometric properties
Validity
Validity, defined as the degree to which an instrument measures the construct(s) it purports to measure (43). For this we evaluated content validity and construct validity. Content validity has been defined as the degree to which the content of the instrument adequately reflects the construct to be measured. We first ascertained this by taking quality indicators from an existing, validated quality assessment instrument (10,11). Next, we discussed the different subjects of the instrument...
Development and validation of a new instrument

565

with nine academic GPs separately, each of them selected because they were expert in the specific subject of the guideline and involved in its development. Adjustments and extensions of the instrument on their field of interest were extensively discussed with them in one-to-one sessions. Subsequently, they were asked for their final approval. Construct validity has been defined as the degree to which the scores of the instrument are consistent with hypotheses (for instance, with regard to relationships to scores of other instruments) based on the assumption that the instrument validly measures the construct to be measured [43]. We compared GP trainees’ guideline adherence at baseline (mean) to the mean baseline score on the national GP knowledge test (LHK), using the Spearman rank correlation coefficient. The LHK is a test measuring knowledge of the professional guidelines. It consists of 120 multiple-choice questions about the national GP CPGs and is used to monitor clinical knowledge development of all GP trainees twice a year [44]. For each GP trainee, guideline adherence was assessed as the number of times a GP trainee adhered to a guideline (i.e. scored +1) as a proportion of all reasons for encounter in the patient consultations assessed with the instrument as seen by that trainee.

Reliability
Reliability has been defined as the degree to produce stable and consistent results [43]. For reliability, we measured the extent to which the two independent raters (both last-year medical students) produced the same results on the same data (intrarater reliability) [43]. We calculated the absolute intraclass correlation coefficient (ICCwhole (2,1)) for three decisions (45,46). First decision was the inclusion (or exclusion) of a patient consultation after the first selection, that was based on ICPC codes. Second was the choice of the assessment item to score a management decision, since for some disorders the choice of assessment item depends on age and the (exact) diagnosis, as interpreted by the rater. For example, there are two asthma CPGs (adults versus children) and three for knee complaints (traumatic versus non-traumatic, adults versus children) (17,18,35,36,47). Third was the allocated guideline adherence score to a management decision on diagnosis, therapy or referral. ICC values were interpreted as: excellent reliability ≥ 0.8, moderate reliability is 0.61–79 and questionable reliability < 0.60 (45).

Feasibility
Feasibility was defined as the viability to assess guideline adherence in a substantial number of patient consultations in a short amount of time per consultation. For feasibility, we first calculated the coverage of the instrument, defined as the number of reasons for encounter in patient consultations presented to GP trainees that were scored, divided by the total number of reasons for encounter in these consultations. In addition, we assessed the time needed to assess patient consultations (in minutes).

Data analyses
Statistical analyses were performed using Statistical Package for the Social Sciences version 20.0 (SPSS Inc., Chicago, IL) and SAS version 9.2 (SAS Institute Inc., Cary, NC). A P-value of ≤0.05 was considered to be statistically significant.

Ethical approval
This study design was assessed by the UMC Utrecht Ethics Committee and regarded as non-eligible for full informed consent. However, we obtained informed consent from the GP trainees for the use of the data from the logs. Patient data were reported anonymously.

Results
Validation process
Data collection
We used data from 12,106 patient consultations, with 12,587 different reasons for encounter, performed by 76 GP trainees at two measurements in the third year of the postgraduate GP training programme. In line with the gender distribution of the GP trainees in the Netherlands, trainees in our study were predominantly female (72%), and the median age was 31 years (interquartile range (IQR) 5 [13,37]). Most (95%) trainees did a hospital internship before starting the GP vocational training programme for a median time period of 20 months (IQR 19). Two (3%) trainees had a PhD. Two-thirds (68%) of the trainees worked full-time during the GP training. The most frequent reasons for encounter presented to the GP trainees were skin problems (ICPC category S, 19%), musculoskeletal disorders (L, 18%) and respiratory complaints (R, 17%). Based on predefined ICPC codes (Table 2), we selected 4203 patient consultations (33% of total) for assessment. We then excluded 1121 patient consultations (9% of total) based on the exclusion criteria, resulting in a final selection of 3082 patient consultations (24% of total) with 8083 different management decisions; 2416 diagnostic (19% of total), 2585 therapeutic (21% of total) and 3082 referral decisions (24% of total).

Psychometric properties
Validity
All nine academic GP experts considered the assessment items on their subject in the instrument to be relevant for daily clinical practice and an adequate reflection of the construct to be measured. The evaluation of construct validity showed that the mean baseline score of the 76 GP trainees on guideline adherence was significantly correlated to their baseline result on the LHK (r 0.33, P 0.004).

Reliability
The interrater reliability on inclusion of a patient consultation for assessment (N = 3010) and on the choice of assessment item to score a management decision (N = 2939) was excellent with, an ICC of 0.80 [P < 0.001, 95% confidence interval (CI) 0.79–0.81] and 1.00 (P < 0.001, 95% CI 1.00–1.00), respectively. Disagreement on inclusion occurred in 8% of patient consultations. Disagreement on the choice of assessment item for a management decision was rare (N = 71, 2% of total) and mainly occurred when the reason for encounter was related to knee problems (N = 53, 75% of disagreements) where the assessment items in the instrument are based on three different guidelines (35,46,47). The interrater reliability for the allocated guideline adherence scores (−1, 0 or +1) to all assessed management decisions (N = 8083) was excellent with an ICC of 1.00 (P < 0.001, 95% CI 1.00–1.00). For diagnostic (N = 2416), therapeutic (N = 2585) and referral (N = 3082) decisions, the interrater reliabilities were moderate with ICCs of 0.79 (P < 0.001, 95% CI 0.78–0.81), 0.79 (P < 0.001, 95% CI 0.77–0.80) and 0.64 (P < 0.001, 95% CI 0.62–0.66), respectively.

Feasibility
The instrument we developed covered 24% (N = 3082) of all reasons for encounter (N = 12,587), presented during the consultations. The mean and median time for both raters to assess patient consultations was around 1 minute; 1.26 (SD 0.5) and 1.04 (SD 0.2) minutes, range 1–7 minutes.
Table 3. Guideline adherence scores in management decisions of 76 third-year GP trainees

<table>
<thead>
<tr>
<th>Score</th>
<th>Diagnostic decisions</th>
<th>Therapeutic decisions</th>
<th>Referral decisions</th>
<th>All decisions</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>108 (4)</td>
<td>311 (12)</td>
<td>240 (1)</td>
<td>443 (5)</td>
</tr>
<tr>
<td>0</td>
<td>378 (16)</td>
<td>64 (24)</td>
<td>299 (10)</td>
<td>1288 (16)</td>
</tr>
<tr>
<td>1</td>
<td>1930 (80)</td>
<td>1663 (64)</td>
<td>2759 (90)</td>
<td>6352 (79)</td>
</tr>
<tr>
<td>Total</td>
<td>2416 (100)</td>
<td>2585 (100)</td>
<td>3082 (100)</td>
<td>8083 (100)</td>
</tr>
</tbody>
</table>

(-1) not in accordance with the guideline and no reason mentioned; (0) debatable if it is in accordance with the guideline due to insufficient or contradicting information; (+1) in accordance with the guideline or a conscious deviation from it.

Guideline adherence scores

In Table 3, the distribution of guideline adherence scores is shown. For the vast majority of the decisions in clinical management we assessed (N = 8083, 79%), GP trainees adhered to the guidelines, especially for the diagnostic decisions (N = 1930, 80%) and decisions on referral (N = 2759, 90%). In 8% of patient consultations (N = 250), trainees appropriately motivated why they deviated from the recommendations in the guideline, with as primary reason patient's preferences (51%). For only a minority of the management decisions (N = 443, 5%), GP trainees did not adhere to the guidelines, where for the remaining 16% it was debatable if GP trainees adhered to the guidelines due to insufficient or contradicting information. Information was missing in 24% of these patient consultations, mainly regarding the medical history (21%) or background information (e.g. comorbidity, allergies, etc.) on the patient (19%). Ten GP trainees (7% of 136) adhered to the guidelines on all diagnostic decisions (with a median of 7 patient consultations per trainee) and 38 trainees (28%) on all referral decisions (with a median of 16 patient consultations per trainee). In 64% of the therapeutic decisions (N = 2585), GP trainees adhered to the guidelines. In total, 58% of the 3082 patient consultations was completely in accordance with the guidelines.

Discussion

Summary

To our knowledge, this is the first instrument assessing guideline adherence of GP trainees in clinical practice, also allowing motivated deviation in individual patients. The instrument enables assessment of three relevant aspects of clinical management decisions (diagnosis, therapy and referral) for a wide variety of disorders, covering one in four GP trainees’ patient consultations. The instrument was demonstrated to be valid, reliable and feasible to use, and facilitates the monitoring of guideline adherence as one of the key indicators of EBM behaviour at the actual point of care during the GP postgraduate training programme.

Strengths and limitations

A major strength is the ability of the instrument to use data from patient consultations of the trainee for a wide range of disorders. The instrument is based on guidelines that are widely accepted and applied in daily practice (8). We validated the instrument in a large number of patient consultations, assessing the recommended properties (validity, reliability and feasibility) (5,7).

We do realize that optimal EBM behaviour in primary care extends beyond guideline adherence alone. Personalized care means taking the best decisions in the perspective of the patient. However, we think that guideline adherence is an important aspect of EBM behaviour, as guidelines are created as a result of a combination of research evidence and clinical expertise, and, more importantly, because we also allowed for deviations from guideline recommendations when appropriately motivated, for example, based on patient preferences (2,8,48).

The method we used to collect patient data for testing the instrument may have led to incomplete data, as we were highly dependent on the way GP trainees recorded their patient consultations. Previous studies show that only one-third of actions undertaken in a patient consultation are recorded with large differences between different parts of the consultations (49). Misclassification of allocated guideline adherence scores may have occurred when trainees did not record reasons to deviate from a guideline. However, with the growing policy needs and demands with respect to political, judicial and insurance aspects in health care, it is expected that patient records will become more comprehensive, (partly) resolving this limitation (50). Moreover, as only one-fifth of all the decisions were scored as ‘debatable whether in accordance with the guideline due to insufficient or contradicting information’ (score 0), we think dependency on missing data was not an important issue in our instrument. For this reason, we expect that our instrument could be used among GPs as well, even though they (as more experienced doctors) are expected to record less and therefore score worse (51). It can be argued that the instrument covers only 24% of all reasons for encounter. Prior research from 1996 (when only 60 national guidelines were available) showed that at that time 27% of all diseases encountered in general practice were covered by the guidelines (52). Extrapolated to the present number of 92 guidelines, the instrument’s coverage could be regarded as relatively good, as it addresses around 40% of all reasons for encounters. Finally, we do realize that the LHK is not the optimal indicator of professional development, as it basically measures theoretical knowledge of professional guidelines (44). However, as no better alternatives are available and as it is widely used in GP training, we think it provides an acceptable benchmark of progress during GP training.

Comparison with existing literature

Previously developed instruments that assess EBM behaviour do not assess guideline adherence, and most of them use indirect methods (i.e. through questionnaires, interviews and focus groups) and not clinical practice data (5). Adherence to guidelines has been assessed before in studies among GPs but has not been used for this purpose before (10,15,53). Compared to these previous studies, where motivated deviation in individual patients was not taken into account, guideline adherence seems comparable, with a slightly better performance among our GP trainees (10,15,53). The main reason trainees mentioned to deviate from a guideline was the patient’s preference, which aligns perfectly with the definition of EBM, as also argued by Greenhalg et al. (54).

Implications for future

The instrument we developed to measure guideline adherence in primary care practice is valid, reliable and feasible. It could be used to monitor clinical performance during professional training (9). Moreover, trainees would be able to reflect on EBM practice behaviour by evaluating their patient consultations using the instrument and discussing the results. When patient data can be derived from EMRs automatically, more (background) information will be...
available (such as patient characteristics, co-morbidity, etc.) enabling optimization of the assessment process. (55) As the use of guidelines is well-established and widely accepted among different specialties and in different countries nowadays (48,56), the concept we used can be applied to other health care settings as well. Our instrument could serve as an example to create an instrument measuring guideline adherence as an indicator of EBM behaviour in clinical practice in a valid, reliable and feasible way.

Acknowledgements

Imke Heerese: As a medical student, she helped us with the development of the first version of the instrument (of 19 adjusted quality indicators) and assessed data as collected by GP trainees with this instrument, as part of a pilot study. Yvonne Spoormons, Lisanne Welink, Jeroen van Duijn: As medical students, they independently scored guideline adherence using the final instrument. All academic experts (Guy Rutten, Alfred Sachs, Theo Verheij, Roger Damoiseaux, Lidewij Broekhuizen, Frans Rutten, Kees Gorter): As academic GPs, they were involved in the development of the instrument, as described in the Methods section. They helped us in modifying quality indicators to assessment items on the clinical topic of their expertise.

Declaration

Funding: none.
Ethical approval: ethical approval has waived.
Conflict of interest: none.

References

40. Van Balen FAM, Rovers MM, Eekhof JAH et al. NHG-Standaard Otitis media met effusie (tweede herziening). In: Wiersma T, Goudswaard AN
41. Rooijackers-Lemmens E, van Balen FAM, Opstelten W, Wiersma Tj. NHG-
staandaard Otitis externa. Eerste herziening. In: Wiersma Tj, Goudswaard
42. Rietveld RP, Cleveringa JP, Blom GH et al. NHG-Standaard Het rode oog
43. Mokkink LB, Terwee CB, Patrick DL et al. The COSMIN study reached
international consensus on taxonomy, terminology, and definitions of
measurement properties for health-related patient-reported outcomes. J
Clin Epidemiol 2010; 63: 737–45.
44. Van Leeuwen YD, Pollemans MC, Moll SS et al. The Dutch knowledge test for
45. Richman J, Makrides L, Prince B. Research methodology and applied sta-
46. Terwee CB, Bot SD, de Boer MR et al. Quality criteria were proposed for
measurement properties of health status questionnaires. J Clin Epidemiol
2007; 60: 34–42.
47. Belo JN, Berg HF, Klein Ilkink AJ et al. NHG-Standaard Traumatische
48. Ansari S, Rashidian A. Guidelines for guidelines: are they up to the task?
A comparative assessment of clinical practice guideline development
49. Rethans JJ, Martin E, Metsemakers J. To what extent do clinical notes by
general practitioners reflect actual medical performance? A study using
50. Koerselman GF, Korzec A. [A proposed checklist when deviating from
52. Wiersma TJ, Rutten GEHM. Standaarden en epidemiologie anno 1996.
In: Thomas S, Geijer RMM, Laan van der JR, Wiersma TJ (eds). NHG-
standaarden voor de huisarts II. Utrecht, The Netherlands: Bunge,
1996.
53. van den Berg MJ, de Bakker DH, Spreeuwenberg P et al. Labour intensity
of guidelines may have a greater effect on adherence than GPs’ workload.
54. Greenhalgh T, Howick J, Maskrey N; Evidence Based Medicine Renais-
348: g3725.
55. van Berkestijn LG, Kastein MR, Lodder A, de Melker RA, Bartelink ML.
How do we compare with our colleagues? Quality of general practitioner
performance in consultations for non-acute abdominal complaints. Int J
Qual Health Care 1999; 11: 475–86.
56. Wolf JS Jr, Hubbard H, Faraday MM, Forrest JB. Clinical practice guide-
lines to inform evidence-based clinical practice. World J Urol 2011; 29:
303–9.