European guidelines for quality assurance in colorectal cancer screening and diagnosis. First Edition
Principles of evidence assessment and methods for reaching recommendations

Co-Funded by
the Health Programme
of the European Union

Authors   S. Minozzi[1], P. Armaroli[1], N. Segnan[1]

Institution   CPO Piemonte (Piedmont Centre for Cancer Prevention) and S. Giovanni Hospital, Turin, Italy

Corresponding author
S. Minozzi
CPO Piemonte (Piedmont
Centre for Cancer Prevention)
and S. Giovanni Hospital
Via San Francesco da Paola 31
10123 Turin
Italy
minozzi.silvia@gmail.com
qas@iarc.fr

Multidisciplinary, evidence-based guidelines for quality assurance in colorectal cancer screening and diagnosis have been developed by experts in a project coordinated by the International Agency for Research on Cancer. The full guideline document covers the entire process of population-based screening. It consists of 10 chapters and over 250 recommendations, graded according to the strength of the recommendation and the supporting evidence. The 450-page guidelines and the extensive evidence base have been published by the European Commission. The principles of evidence assessment and methods for reaching recommendations are presented here to promote international discussion and collaboration by making the principles and methods used in developing the guidelines known to a wider professional and scientific community. Following this methodology in the future updating of the guidelines has the potential to enhance the control of colorectal cancer through improvement in the quality and effectiveness of the screening process, including multidisciplinary diagnosis and management of the disease.

Background
▼

According to the most recent estimates by the International Agency for Research on Cancer [3] colorectal cancer (CRC) is the most common cancer in Europe with 432 000 new cases in men and women reported annually. It is the second most common cause of cancer deaths in Europe with 212 000 deaths reported in 2008. Worldwide CRC ranks third in incidence and fourth in mortality with an estimated 1.2 million cases and 0.6 million deaths annually. The European Union (EU) recommends population-based screening for breast, cervical and colorectal cancer using evidence-based tests with quality assurance of the entire screening process including diagnosis and management of patients with screen-detected lesions [1]. The EU policy takes into account the principles of cancer screening developed by the World Health Organization [16] and the extensive experience in the EU in piloting and implementing population-based cancer screening programmes [12]. Screening is an important tool in cancer control in countries with a significant burden of CRC, provided the screening services are high quality [13]. The presently reported multidisciplinary, evidence-based guidelines for quality assurance in colorectal cancer screening and diagnosis have been developed by experts and published by the EU [10].

Results
▼

The principles of evidence assessment and the methods for developing the recommendations presented in the Guidelines are described below. The contribution of the Literature Group was crucial to the feasibility of this resource-intensive process. It included assistance to the chapter authors in defining relevant clinical questions of key importance.

The clinical questions for which evidence was collected by the Literature Group and the results of the literature search and analysis conducted by the group are presented in Appendix 1 to the Guidelines [6]. The appendix is only available in electronic format due to the extensive size of the records that correspond to approximately 1000 printed pages.

Introduction
▼

The evidence-based process for development of the recommendations in the first edition of the European guidelines was established at the outset

of the project in 2006 by an editorial board with extensive experience in development of best practice guidelines, in evaluation of strategies for colorectal cancer (CRC) screening and in programme management. In 2007 the editorial board drafted an initial comprehensive outline of the Guidelines and recruited a multidisciplinary group of experts in colorectal cancer screening and diagnosis across the European Union to collaborate in revising the outline and drafting the chapters, including guiding principles and recommendations. Additional scientific support was provided by a Literature Group consisting of epidemiologists with special expertise in the field of CRC and in performing systematic literature reviews.

The expert Literature Group provided technical and scientific support to the authors and editors in searching the relevant literature, assessing the methodological quality of retrieved studies, defining a grading system of the level of evidence and strength of the recommendations, and preparing evidence tables and summary documents for over 500 references identified through systematic reviews of the literature according to the priorities and procedures agreed with the editorial board and the authors.

The Literature Group was coordinated by N. Segnan at the Unit of Cancer Epidemiology, Department of Oncology of the Piedmont Centre for Cancer Prevention (CPO Piemonte) and S. Giovanni University Hospital, Turin, Italy, and was lead by S. Minozzi at the same institution. Other members of the Literature Group were based at the CPO in Turin and at the Oxford University Cancer Screening Research Unit, Cancer Epidemiology Unit, Oxford, United Kingdom. Additional scientific and technical support was provided by the International Agency for Research on Cancer, Quality Assurance Group, Section of Early Detection and Prevention, Lyon, France.

Definition of clinical questions

In multidisciplinary workshops conducted in 2007 and 2008 the chapter authors met with the editorial board and the Literature Group. At these meetings, the table of contents of the Guidelines was repeatedly revised and the methodology of evidence-based guideline development, including the process of identifying and evaluating the relevant evidence for each chapter based on the topics in the revised outline was agreed with the authors. Subgroups of authors responsible for each chapter also worked individually with members of the Literature Group to develop clinically relevant questions based on the revised chapter outlines, and the results for each chapter were subsequently discussed with the entire group of authors and editors and the Literature Group in plenary workshop sessions in order to ensure a common methodological approach and to reach a consensus on questions of key importance requiring the support of the Literature Group in order to identify and assess the relevant evidence. This collaborative, multidisciplinary approach remained a guiding principle throughout the entire process up to completion of drafting and editing of the Guideline chapters.

The clinical questions initially formulated by the authors of each chapter and subsequently agreed with the editorial board and the other authors were developed according to the PICOS method [4,8,9] modified slightly to take into account the aim of screening to lower the burden of the disease in the population:

P: patients/population characteristics
I: experimental **intervention** on which the question is focused
C: comparison intervention/control/reference group
O: outcome measure relevant for the clinical question
S: study design on which to base the evidence search

The extensive list of initial clinical questions was reduced to a feasible number, by prioritising questions of key importance for each chapter. In total, 113 clinical questions were prioritised. The PICOS components of each prioritised question were subsequently used by the Literature Group to define specific key words that were then employed in comprehensive bibliographic searches. The results of these activities were reported back to the authors and editors in subsequent workshops and electronically. This enabled the editors and authors to provide continuous professional and scientific support to the process of identifying and analysing the relevant evidence.

Bibliographic review

The Literature Group performed bibliographic searches on Medline, and in many cases also on Embase and The Cochrane Library using MeSH terms and free text words. Most searches were limited to the years 2000 to 2008 or were conducted without date restrictions if the authors or editors who were experts in the field knew that there were relevant articles published before 2000. Published articles suggested by the authors and not retrieved by a systematic search were also considered. Only scientific publications in English, Italian, French and Spanish were included. Priority was given to recently published, systematic reviews or clinical guidelines. If systematic reviews of high methodological quality were retrieved, the search for primary studies was limited to those published after the last search date of the most recently published systematic review (i.e. if the systematic review had searched primary studies until February 2006, primary studies published after February 2006 were sought). If no systematic reviews were found, a search for primary studies published since 2000 was performed.

In selected cases references not identified by the above process were included in the evidence base, i.e. when authors of the chapters found relevant articles published after 2008 during the period when chapter manuscripts were drafted and revised prior to publication. The criteria for relevance were: articles concerning new and emerging technologies where research is growing rapidly, high quality and updated systematic reviews, and large trials that make a significant contribution to the robustness of the results or allow upgrading of the level of evidence.

Inclusion criteria

The inclusion criteria applied by the Literature Group were based on the highest level of available evidence, taking into account study design. For primary studies, for each kind of question (e.g., effectiveness, diagnostic accuracy, acceptability and compliance) a hierarchy of the study designs and inclusion/exclusion criteria was developed by the epidemiologists in the Literature Group. For example, for effectiveness studies randomised controlled trials (RCT) were initially searched for. If RCTs were retrieved, no other types of study design were considered. If no, or only a few and/or small RCTs were retrieved, quasi-experimental studies were considered. If no quasi-experimental studies were found, prospective or retrospective cohort and case-control studies were considered. If studies with none of the above designs were retrieved, cross-sectional studies and case series were included. For diagnostic accuracy questions, cross-sectional studies with verification by reference standard were considered as the best source of evidence.

Quality assessment

The methodological quality of the publications retrieved by the Literature Group was assessed using the following criteria obtained from published and validated check lists.

Systematic reviews – quorum checklist

A validated checklist for evaluating the manner in which systematic reviews have been conducted was not available when the methods for the present EU Guidelines were established. Therefore the QUOROM checklist that assesses the quality of reporting was used as a proxy to assess the quality of conduct of systematic reviews. This approach reflects the view that the quality of reporting can be used as a criterion for the quality of the process of preparing a systematic review [7].

Randomised Controlled Trials

Randomised controlled trials were assessed using the following criteria suggested in the Cochrane Handbook [5] and by the Cochrane Effective Practice and Organisation of Care Review Group [2]:

- ▶ Unit of allocation (i.e. who or what was allocated to study groups: individuals or clusters);
- ▶ Unit of analysis (i.e. results analysed as events at the level of individuals or clusters);
- ▶ If unit of allocation and unit of analysis differ, was cluster analysis performed?
- ▶ Protection against selection bias (adequate sequence generation and allocation concealment);
- ▶ Protection against performance bias (blinding of providers);
- ▶ Protection against contamination (blinding of participants);
- ▶ Protection against attrition bias (intention to treat analysis, few lost at follow up balanced between groups); and
- ▶ Protection against detection bias (blinding of participants and outcome assessors).

Observational studies: cohort studies and case control studies

Observational studies were evaluated using the following criteria of the Newcastle-Ottawa Scale (for recent overview see: [14]

- ▶ Case control studies:
  - ▷ Adequate definition of the cases;
  - ▷ Representativeness of the cases;
  - ▷ Selection source of controls;
  - ▷ Definition of controls;
  - ▷ Comparability of cases and controls on the basis of the design or analysis;
  - ▷ Method of exposure assessment;
  - ▷ Same method of ascertainment for cases and controls;
  - ▷ Non-Response rate.
- ▶ Cohort studies:
  - ▷ Representativeness of the exposed cohort;
  - ▷ Selection source of the non-exposed cohort;
  - ▷ Method of exposure assessment;
  - ▷ Demonstration that outcome of interest was not present at start of study;
  - ▷ Comparability of cohorts on the basis of the design or analysis;
  - ▷ Method outcome assessment;
  - ▷ Adequacy of follow up of cohorts.

Interrupted time series studies

Studies based on interrupted time series were assessed using the following criteria suggested by the Cochrane Effective Practice and Organisation of Care Review Group (EPOC 2002):

- ▶ Clearly defined point in time when the intervention occurred.
  - ▷ A: Intervention occurred at a clearly defined point in time;
  - ▷ B: NOT CLEAR because not reported in the paper;
  - ▷ C: Intervention did not occur at a clearly defined point in time.
- ▶ At least three data points before and three after the intervention.
  - ▷ A: Three or more data points before and three or more data points recorded after the intervention;
  - ▷ B: NOT CLEAR because not reported in the paper;
  - ▷ C: Less than three data points recorded before, and less than three data points recorded after intervention.
- ▶ Protection against secular changes (the intervention is independent of other changes).
  - ▷ A: Intervention occurred independently of other changes over time;
  - ▷ B: NOT CLEAR because not reported in the paper;
  - ▷ C: Intervention was not independent of other changes over time.
- ▶ Protection against detection bias (intervention unlikely to affect data collection).
  - ▷ A: Intervention unlikely to affect data collection (for example, sources and methods of data collection were the same before and after the intervention);
  - ▷ B: NOT CLEAR because not reported in the paper;
  - ▷ C: Intervention likely to affect data collection (for example, any change in source or method of data collection before vs. after the intervention).
- ▶ Blinded assessment of primary outcome(s).
  - ▷ A: Explicit statement of authors that the primary outcome variables were assessed blindly OR the outcome variables are objective e.g. length of hospital stay, drug levels as assessed by a standardised test;
  - ▷ B: NOT CLEAR if not specified;
  - ▷ C: Outcomes were not assessed blindly.
- ▶ Completeness of data set.
  - ▷ A: Data set covers 80–100% of total number of participants or episodes of care in the study;
  - ▷ B: NOT CLEAR if not specified;
  - ▷ C: Data set covers less than 80% of the total number of participants or episodes of care in the study.

Diagnostic accuracy studies

The criteria used to evaluate diagnostic accuracy studies were obtained from the QUADAS checklist [15]:

- ▶ Study design: diagnostic cross-sectional studies with prospective or retrospective recruitment; case control;
- ▶ Spectrum of patients representative of the individuals who will receive the test in practice;
- ▶ Patients selection criteria clearly described;
- ▶ Verification by reference standard of all or a randomised sample of subjects (absence of verification bias);
- ▶ Execution of the index and comparator tests adequately described;
- ▶ Execution of the reference standard adequately described;

▶ Independent and blind interpretation of index test and reference standard results;
▶ Un-interpretable/intermediate test results reported;
▶ Withdrawals from the study explained.

Clinical guidelines

The quality of clinical guidelines evaluated by the Literature Group was assessed using the following most relevant criteria derived from the COGS checklist [11]:
▶ Description of the clinical specialisation of the members of the panel of guideline authors;
▶ Search strategy described (databases, years covered, any language restriction);
▶ Inclusion criteria of primary studies stated;
▶ Method used to analyse and synthesise the evidence and to reach the consensus among the panellists to elaborate the recommendation described;
▶ Presence of a grading of level of evidence and/or of the strength of the recommendation; and
▶ Presence of a complete reference list.

Evidence tables and summary documents

The Literature Group prepared the following documents based on the publications retrieved for each clinical question or group of clinical questions. The documents were subsequently used by the authors in drafting respective chapters:
▶ An evidence table for each retrieved study with the main characteristics of the study (study design, objective of the study, comparisons, participant's characteristics, outcome measures, results, methodological quality, level of evidence);
▶ A summary document with a synthesis of the number, types and characteristics of the retrieved studies, their overall methodological quality, a description of the main methodological flaws, the study results and the conclusions and the overall level of evidence.

The evidence tables and summary documents for each chapter are documented in [6]. Evidence tables were not prepared for: additional publications cited in the background sections of the chapters; pathological and clinical classifications; technical instructions; narrative reviews; editorials and personal communications; and articles published before 2000 and cited by the authors after the systematic search of the literature.

Some articles published between 2000 and 2008 and not retrieved by the systematic search were considered to be relevant by the authors. Those references have therefore been included in the body of evidence in agreement with the editorial board. For these articles, additional evidence tables were prepared after December 2009, but the respective results were not included in the respective summary documents.

Grading system

The key recommendations presented in each chapter of the Guidelines are listed at the front of the respective chapter together with a grading of the evidence on which each recommendation is based, and the strength of the recommendation. Only the highest level of evidence supporting a recommendation is reported. The following grading scales are used:

Level of the evidence
▶ **I:** multiple randomised controlled trials (RCTs) of reasonable sample size, or systematic reviews (SRs) of RCTs
▶ **II:** one RCT of reasonable sample size, or 3 or less RCTs with small sample size
▶ **III:** prospective or retrospective cohort studies or SRs of cohort studies; diagnostic cross section-al accuracy studies
▶ **IV:** retrospective case-control studies or SRs of case-control studies, time-series analyses
▶ **V:** case series; before/after studies without control group, cross sectional surveys
▶ **VI:** expert opinion

Strength of the recommendations

The strength of recommendations was graded according to the following scale:
▶ **A:** intervention strongly recommended for all patients or targeted individuals
▶ **B:** intervention recommended
▶ **C:** intervention to be considered but with uncertainty about its impact
▶ **D:** intervention not recommended
▶ **E:** intervention strongly not recommended

The strength of each key recommendation was determined by the authors of each chapter in agreement with the Guidelines editorial board.

Following the list of key recommendations at the beginning of each chapter, the rationale and the evidence on which the recommendations are based is summarised in the body of the chapter, including the respective levels of evidence.

In a number of chapters, in addition to the key recommendations, fundamental statements (Guiding Principles) defining the aims and scope of the recommendations presented in the chapter are provided at the front of the text. Most of the Guiding Principles are considered to be self-evident. All reflect the consensus of the authors and editors on essential principles of best practice in screening and diagnosis of colorectal cancer. In addition to these principles, additional advisory statements are made in the body of the chapters that are not specifically graded. These statements also represent the consensus of the authors and editors on best practice.

Correspondence between level of evidence and strength of recommendation

This present grading of the strength of recommendations did not require a rigid correspondence with the levels of evidence. For example grade **A** was given to interventions for which there was evidence level **I** (multiple RCTs or SR of RCTs) but also to interventions that could not be assessed by RCTs, (e.g. psychological aspects, the importance of an accurate information to the patients, etc). Grade **B** was given to interventions with lower evidence level (**II** or **III**) but also for interventions with evidence level **I** but with uncertainty about their impact in the population or about practical implementation (e.g. lack of resources for implementation, social barriers, supposed lack of acceptability by the target population). Grade **C** level was given to interventions for which evidence was not available or was of low grade (i.e. **IV**, **V**) or that may not have been considered of high importance for other reasons (i.e. psychological or social aspects). Grades **D** and **E** were assigned to interventions for which there was evidence of no benefit for participants, or for which the harm outweighed the benefits. ● **Table 1**

Levels of evidence	Strength of recommendation					
		A	B	C	D	E
	I	C	C		C	C
	II	Nc	C		C	Nc
	III	Nc	C	C	C	Nc
	IV	Nc	Nc	C	Nc	Nc
	V	Nc	Nc	C	Nc	Nc
	VI	Nc	Nc	C	Nc	Nc

Table 1 Correspondence between level of evidence and strength of recommendations

C: Coherence between the level of evidence and the strength of recommendations
Nc: No coherence between the level of evidence and the strength of recommendations

Method of obtaining consensus between the chapter authors and editors and the internal peer review

Each subgroup of authors responsible for a chapter received all the evidence tables and summary documents relating to the respective clinical questions. The authors drafted each chapter by describing the relevant issues, summarising the evidence, and including recommendations and conclusions. The authors also proposed a grading for the strength of the evidence and the strength of the respective recommendations, based on the results of the literature search and on their clinical experience, as well as any additional pertinent scientific literature that was taken into account with agreement from the editorial board. The draft chapters and the proposed strength of each recommendation were discussed with the editorial board and the authors of all chapters to reach consensus.

External peer review

Chapter drafts were subsequently sent to international experts in their respective fields for external peer review. They were also made available for web consultation with restricted access by experts involved in screening programmes. Comments and criticisms were considered and a final version of the chapters was elaborated. Preliminary and nearly final versions of the Guidelines chapters were prepared and discussed at pan-European network meetings of screening experts, clinicians, advocates, healthcare planners and regulators from all of the EU member states and two EU applicant countries in 2008 and 2009.

Final editing

During 2010, final changes resulting from the network discussion in November 2009 were taken into account by the authors of respective chapters. The consistency of the recommendations between the individual chapters was reviewed by the editorial board and corrections were made where necessary.

Discussion
▼

In the preparation of comprehensive guidelines for quality assurance in colorectal cancer screening and diagnosis an extensive body of scientific evidence has been systematically collected and reviewed by a multidisciplinary group of experts. The editors recognise that the approach to collection of the relevant evidence may have permitted introduction of bias if the authors or editors were not aware of significant publications after 2008 because many of the systematic searches performed by the Literature Group did not include subsequent years. However, the relevant publications of studies published after 2008 that have been cited by the authors to justify recommendations have been evaluated by the Literature Group, and respective evidence tables are inclu-

ded in [6]. In view of the qualifications and experience of the authors and editors and the transparency of the process of guideline development, the editors have concluded that further efforts to limit this potential bias would have little or no impact on the content of the present recommendations.

The editors hope that the approach to evidence-based guideline development adopted for the first edition of the European Guidelines will promote systematic discussion of the evidence base for the Guidelines and that resources will be available in the future to continuously update and expand the current evidence base and the respective documentation.

Conclusions
▼

Principles of evidence assessment and systematic methods for reaching evidence-based standards and recommendations for best practice in colorectal cancer screening and diagnosis have been developed and applied in a collaborative pan-European project. Following this methodology in the future updating of the European Guidelines has the potential to enhance the control of colorectal cancer through improvement in the quality and effectiveness of the screening process, including multidisciplinary diagnosis and management of the disease.

Disclaimer
▼

The views expressed in this document are those of the authors. Neither the European Commission nor any person acting on its behalf can be held responsible for any use that may be made of the information in this document.

Competing interests: No competing interests reported.

Acknowledgements
▼

ment no. 2005317: Development of European Guidelines for Quality Assurance of Colorectal Cancer Screening). Partner institutions: Oxford University Cancer Screening Research Unit, Cancer Epidemiology Unit, University of Oxford, Oxford, United Kingdom; Unit of Cancer Epidemiology, Centre for Cancer Epidemiology and Prevention (CPO) and S. Giovanni University Hospital, Turin, Italy; Public Association for Healthy People, Budapest, Hungary; European Cancer Patient Coalition (ECPC), Utrecht, Netherlands; Quality Assurance Group, Section of Early Detection and Prevention, International Agency for Research on Cancer, Lyon, France.

References

1 Council of the European Union. Council Recommendation of 2 December 2003 on cancer screening (2003/878/EC). Off J Eur Union; 2003: 34 – 38

2 Cochrane Effective Practice and Organisation of Care Group (EPOC). Data Collection Checklist. 2002: http://www.epoc.cochrane.org/en/handsearchers.html

3 *Ferlay J, Shin HR, Bray F* et al. GLOBOCAN 2008 v1.2, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10. [Internet] Lyon, France: International Agency for Research on Cancer; 2010: Available from: http://globocan.iarc.fr Accessed on 05/04/2012

4 *Greenhalgh T.* Why read papers at all? How to read a paper. The basics of evidence-based medicine. BMJ Books; 1997: 1 – 14

5 *Higgins JPT, Altman DG.* Assessing risk of bias in included studies. In: Higgins JPT, Green S (eds.) Cochrane Handbook for Systematic Reviews of Interventions. UK: Wiley-Blackwell; 2008

6 *Minozzi S, Armaroli P, Banzi R* et al. European guidelines for quality assurance in colorectal cancer screening and diagnosis – First edition. Appendix 1: Systematic evidence review 2010: http://bookshop.europa.eu/en/european-guidelines-for-quality-assurance-in-colorectal-cancer-screening-and-diagnosis-pbND3210390/ Accessed 11/2/2012

7 *Moher D, Cook DJ, Eastwood S* et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. Lancet 1999; 354: 1896 – 1900

8 *O'Connor D, Green S, Higgins JPT.* Defining the review question and developing criteria for including studies. In: Higgins JPT, Green S (eds.) Cochrane Handbook for Systematic Reviews of Interventions (Wiley Cochrane Series). UK: Wiley-Blackwell; 2008

9 *Richardson WS, Wilson MC, Nishikawa J* et al. The well-built clinical question: a key to evidence-based decisions. ACP J Club 1995; 123: A12 – A13

10 *Segnan N, Patnick J, von Karsa L* (eds.) European guidelines for quality assurance in colorectal cancer screening and diagnosis – First edition. Luxembourg: European Commission, Publications Office of the European Union; 2010

11 *Shiffman RN, Shekelle P, Overhage JM* et al. Standardized reporting of clinical practice guidelines: a proposal from the Conference on Guideline Standardization. Ann Intern Med 2003; 139: 493 – 498

12 *von Karsa L, Anttila A, Ronco G* et al. Cancer Screening in the European Union. Report on the implementation of the Council Recommendation on Cancer Screening – First Report Luxembourg: European Commission; 2008: http://ec.europa.eu/health/archive/ph_determinants/genetics/documents/cancer_screening.pdf Accessed 11/2/2012

13 *von Karsa L, Lignini TA, Patnick J* et al. The dimensions of the CRC problem. Best Pract Res Clin Gastroenterol 2010; 24: 381 – 396

14 *Wells GA, Shea B, O'Connell D* et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. 2010: http://www.ohri.ca/programs/clinical_epidemiology/oxford.htm

15 *Whiting P, Rutjes AW, Reitsma JB* et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003; 3: 25

16 *Wilson JM, Jungner YG.* Principles and practice of mass screening for disease. World Health Organization; 1968: http://whqlibdoc.who.int/php/WHO_PHP_34.pdf Accessed 11/2/2012