I mproving
Health Care
and
O utcomes

Method for Evaluating Research Guideline Evidence

NSW HEALTH

Improving Health Care and Outcomes

Method for Evaluating Research Guideline Evidence

Jeannine Liddle, Margaret Williamson and Les Irwig

December 1996

ISBN 0 7310 9248 1
State Health Publication No. (CEB) 96-204

Suggested Citation: Liddle J, Williamson M, Irwig L. Method for evaluating research and guideline evidence. NSW Health Department, Sydney, December 1996.

For additional copies of this document contact:

Better Health Centre
162 Blues Point Road
North Sydney NSW 2060
AUSTRALIA
Telephone: 02 391 9010
Fascimile: 02 955 5196

This document is also available on the World Wide Web at
http://www.health.nsw.gov.au/public health

Table of Contents

List of Tables

Acknowledgments

Special thanks to Lyn March and Anne Chamberlain from the Northern Sydney Area Health Service Fractured Neck of Femur health outcomes group who provided ongoing input into the development and piloting of MERGE.

We also acknowledge comments from Deborah Baker, Tien Chey, Glenn Close, Robert Cumming, Michael Frommer, Paul Glasziou, Penny Hawe, David Henderson-Smart, Dianne Kelleher, Louisa Jorm, Hanna Noworytko, Helen Moore, Geoff Morgan, Cait Lonie, Chris Silagy, Marie-Louise Stokes, Bernie Towler and Andrew Wilson.

We would also like to thank members of the NSW Health Department Expert Panel on Diabetes, Guidelines Working Group: Steven Boyages, Lesley Campbell, Ruth Colagiuri, Stephen Colagiuri, Robert Moses and Dennis Yue.

SECTION 1 INTRODUCTION

1.1 What is the method for evaluating research and guideline evidence?

The Centre for Clinical Policy and Practice of the NSW Health Department has developed a health outcomes model which aims to improve people's health by focusing on the outcomes as well as the processes of health care and prevention.

In applying the model to diabetes management, we wanted to develop guidelines for promoting health care practices which lead to the best possible outcomes for people with diabetes. The Method for Evaluating Research and Guideline Evidence (MERGE) sets out an explicit standardised approach to reviewing and incorporating scientific evidence into guidelines. MERGE helps guideline developers comply with principles outlined by the NH&MRC Quality of Care and Health Outcomes Committee (QCHOC):

- Clinical practice guidelines should be based on the best available evidence.
- The method used to synthesise the evidence should be the strongest applicable.
- Guidelines should contain a statement concerning the strength of recommendations.[1]

MERGE is divided into five main sections:

Section 1 **Introduction** – Puts MERGE in context! This section describes how MERGE was developed and how to use it.

Section 2 **Study Checklists** – Sets out criteria in question form in five different checklists. These checklists allow reviewers to assess the quality of studies and reviews.

Section 3 **Summarising evidence from individual studies** – Gives some tips on how to do this using the summary of evidence format.

Section 4 **Intervention guidelines and recommendations checklist** – Sets out criteria in question form to allow reviewers to assess whether guidelines or recommendations are valid and likely to benefit a population.

Section 5 **Supplementary notes** – Contains explanatory notes and definitions, references, bibliography and an example of the Summary of Evidence Format. The bibliography contains publications used by authors for developing the checklists and classifying levels of evidence.

The development of MERGE is ongoing.
Comments on MERGE are welcomed and can be forwarded to:
Centre for Clinical Policy and Practice
NSW Health Department
LMB 961, NORTH SYDNEY, NSW 2059
AUSTRALIA.

1.2 Why produce a *new* evaluation method?

Existing checklists[2] for evaluating the quality of scientific evidence focused on randomised controlled trials (RCTs), were often ill-defined, required high level epidemiological skills to use and needed a significant investment of time and resources.

Seven guidelines had been recommended for the clinical management of diabetes by an expert working group and before their implementation, we wished to ensure that the interventions recommended by those guidelines were supported by the best evidence available.

We needed a way of evaluating the scientific literature which:

- was relatively quick, given our limited time and resources.
- was explicit and standardised, to ensure consistency by the users.
- could be used for clinical and public health interventions where RCTs were not available or could not be done.

1.3 How was MERGE developed?

MERGE was developed through consultation with other epidemiologists working in Australia, the Cochrane Collaboration and clinicians on the NSW Health Department Expert Panel on Diabetes Guidelines Working Group. Each of these groups provided information e.g. main areas of potential bias that can distort study results, existing checklists in use, content issues relating to guideline development. We also reviewed the literature relating to 'bias', 'study design' and guideline development. Early drafts of this document were piloted in the clinical management of diabetes and the prevention, management and rehabilitation of fractured neck of femur.

An initial document: *Evaluation checklist for evidence-based guidelines* (May 1995)[3] was widely circulated for comment. Comments and results from inter-rater reliability of the draft checklists indicated areas which required further development, rewording or further explanation. This led to the development of a combined checklist to assess randomised controlled trials and studies, non-randomised controlled studies, cohorts, case-control studies, before and after studies as well as a new checklist for interrupted time series studies. MERGE was further reviewed to incorporate work done by the Cochrane Collaboration on Effective Professional Practice (CCEPP)[4] and the University of York NHS Centre for Reviews and Dissemination.[5]

1.4 When to use MERGE

The two main reasons for using MERGE are:

- to evaluate the quality of evidence from individual studies.
- to evaluate the validity of intervention guidelines and recommendations.

Other applications:

- to suggest how to incorporate evidence from individual studies into a review of evidence.
- to evaluate the quality of scientific evidence for clinical interventions.
- to evaluate the quality of evidence for proposed public health strategies which might be developed at a local or state-wide level.
- to evaluate the quality of studies submitted for publication in peer reviewed journals, and
- to ensure important methodological aspects of study design and performance are reported in journal articles.

Table 1 indicates which section to use, depending on *your* reasons for using MERGE.

Table 1: When to use MERGE

Reason for using MERGE	Document being reviewed	Use Checklist	Outcome of using Checklist
Reviewing quality of studies for clinical or public health interventions	Review article, meta-analysis	Checklist 1: Reviews of the effect of interventions	Assessment of the quality of the review using explicit criteria
	Randomised controlled trial or study	Checklist 2: Studies assessing the effect of interventions	Assessment of the quality of the study using explicit criteria
	Non randomised controlled trial or study		
	Cohort study		
	Case control study		
	Before and after study		
	Interrupted time series study	Checklist 3: Interrupted time series studies assessing the effect of interventions	Assessment of the quality of the ITS study using explicit criteria
Reviewing quality of studies to assess risk factors	Study assessing risk factors	Checklist 4: Studies assessing risk factors	Assessment of the quality of the study assessing risk factors using explicit criteria
Reviewing quality of studies to assess diagnostic accuracy	Study assessing diagnostic accuracy	Checklist 5: Studies assessing diagnostic accuracy	Assessment of the quality of the study assessing diagnostic accuracy using explicit criteria
Summarising evidence from studies into a review of evidence	Completed study checklists	Summary of evidence format	Summary of the evidence from the study checklists is in a standardised format
Reviewing existing guidelines and recommendations or guidelines and recommendations being developed	Intervention Guideline or Recommendation	Checklist 6: Intervention guidelines and recommendations	Assessment of the validity of evidence supporting the intervention guidelines and recommendations (following the review of study quality using the study checklists). Identification of • Benefits and harms of intervention guidelines and recommendations • Areas for further research, applicability of evidence to target population, implications for policy development.

1.5 How MERGE fits into the framework for developing and evaluating guidelines

The following framework outlines the guideline process recommended by QCHOC[1] and highlights the role of MERGE in this process.

Stage 1. Determine the need for and scope of the guidelines or health recommendation

As well as assessing the size of the health problem to be addressed by the guidelines, the group should identify and evaluate any existing guidelines which may address the problem.

Stage 2. Developing guidelines

2.1. *Convene a Multidisciplinary Expert (Guideline) Group*

2.2 *Define the scope of guidelines including the interventions to be assessed and the intended users.*

2.3 *Identify health outcomes*

2.4 *Review the scientific evidence*

MERGE can be used as an alternative or an adjunct to the Cochrane Collaboration Handbook[6] especially in assessing non-RCTs.

2.5 *Formulate the guidelines*

MERGE can assist in defining the likely outcomes of each intervention, in comparing the different interventions based on the strength of the evidence and ensuring that the benefits and risks are explicitly presented and considered.

2.6 *Formulate an implementation strategy*

2.7 *Formulate an evaluation plan*

2.8 *Produce the guidelines and report on the guideline development process*

2.9 *Assess the characteristics of the guidelines*

MERGE can be used to assess whether the guideline is evidence-based. More detailed and sophisticated instruments such as that developed by the US Institute of Medicine[7] are available for more comprehensive assessments of guideline validity.

2.10 *Consult other groups*

2.11 *Pilot test*

2.12 *Re-draft the guideline document*

Stage 3 Implementing the guidelines

Guidelines are not self-implementing and require specific planned implementation strategies.[8,9,10]

Stage 4 Evaluating and updating the guidelines

4.1 *Implement the evaluation plan*

4.2 *Report on the evaluation*

4.3 *Review and update the guidelines regularly*

MERGE may be used as in Stage 2 to assess any new scientific evidence, assist in defining the likely outcomes of new interventions and compare the new intervention to those recommended by the current guidelines, based on the strength of the evidence and the balance of risks and benefits.

1.6 Using MERGE to review the evidence for guidelines and recommendations

When MERGE is used to review the scientific evidence for guidelines or recommendations, we suggest it form part of a structured, comprehensive strategy. The components of such a strategy are:

- **Establish a review group** which includes individuals who will carry out the review, content experts and an epidemiologist.

- **Check if a review has already been completed.** Contact the Cochrane Collaboration, an international network of people, systematically and rigorously reviewing evidence for a wide range of interventions. Cochrane review groups may already be working on your clinical interest area and have published results - this would save unnecessary duplication of effort. Cochrane review groups are committed to handsearching the literature and updating their reviews as new evidence emerges. If a Cochrane review has been completed, the quality of the review would be assessed using MERGE and the evidence incorporated into the relevant guideline or recommendation.

  The Australasian Cochrane Centre can be contacted at:

  Australasian Cochrane Centre
  Flinders Medical Centre
  Bedford Park, SA 5042
  Australia
  Tel: +61 (8) 204 5255
  Fax: +61 (8) 276 3305
  Email: cochrane@flinders.edu.au

- **Agree on a systematic approach to locating the scientific literature, the details of the review procedure, and how the review will be summarised and incorporated into the guideline or recommendation.**

  One suggested approach is:

  Step 1 **Select reviewer(s) and agree on details of the review procedure.**
  Select reviewers who have previous training or experience in critical appraisal of the medical literature. We recommend using at least two reviewers.

  Where several people are involved in reviewing the evidence, it is important for the group to decide how they will use MERGE, what criteria or terms in MERGE need clarification and how they will resolve differences in opinion on the quality of particular studies. Ideally, the group should try out MERGE on a few studies first and work through these issues before proceeding with the review. The group should also decide to what extent they will search the literature and make this explicit in the summary of evidence. The depth of the review will depend on the time and resources available. Once the review is under way, we recommend that each reviewer read the methods section of each study at least twice. At times, the information needed by MERGE is not obvious at the first read. The group should meet regularly during the review to discuss problems, any differences in opinion on the quality of particular studies, issues relating to evaluation criteria and defining levels of evidence.

Step 2 **Specify the objective of the review of evidence.**

Specifying the objective helps ensure that the volume of evidence is manageable and that irrelevant and unnecessary literature searches are avoided. Experts and intended users of the guidelines or recommendations can advise what is important to know and what they would like clarified through the evidence. Consider harms as well as benefits of an intervention.

Step 3 **Identify strategies to locate the full range of evidence including unpublished results and work in progress.**

Evidence will be identified from electronic databases such as MEDLINE and CINAHL using a variety of search strategies (meta-analysis, randomised, clinical trial, keywords from content area) as well as from content experts and articles referred to in these sources and other bibliographies. This step may generate a large list of abstracts needing further classification in Step 4.

Step 4 **Classify the literature according to general purpose and study type.**

The focus here should be on the methods section of the abstract rather than the results, because a knowledge of the results might bias the assessor in deciding whether to retrieve the full article in Step 5.

Step 5 **Retrieve the full version of evidence available.**

This step entails deciding which articles to retrieve in full. It is important to record these decisions. For example, reviewers may decide beforehand only to retrieve randomised controlled trials.

Step 6 **Assess the quality of the evidence.**

This takes into account the extent to which systematic errors (bias) have been prevented in study design and execution. Bias may lead to an over or underestimation of the 'true' effectiveness of an intervention. MERGE checklists are used in this step.

Step 7 **Quantify the strength of the evidence.**

Where possible the strength of evidence should be quantified using meta-analysis techniques, summarising the results into a single point estimate with confidence intervals for each benefit and harm. Results from subgroup analysis can also be summarised where relevant. Where results cannot be combined to quantify the strength of the evidence, the reason for this should be explained and the results of each study should be presented separately.

Step 8 **Express the evidence in a standard way.**

Once the evidence from the studies has been reviewed, it should be amalgamated into a summary which includes the following aspects:

- *Summary* of major points of the evidence and rating of the strength of evidence
- *Search strategy* used
- *Quality of evidence* considering epidemiological and content issues
- *Results* summarising the benefits and harms of the intervention
- *Conclusion* based on the strength of the association between the intervention and the harms and benefits
- *References*.

For more details of the summary format see Section 3.2.

These steps were followed to review the evidence for effective interventions in the clinical management of diabetes.[11]

1.7 What happens next?

What happens next is beyond the scope of this document. The next steps involve the *systematic* implementation of the guidelines or recommendations, including an assessment of the feasibility and cost of their implementation and the systematic evaluation of their outcomes. This is indeed a challenge but essential if the evidence is to be applied in practice.

SECTION 2 STUDY CHECKLISTS

There are **five study checklists** for evaluating the quality of studies depending on their study type and study purpose. The five checklists are for:

- Checklist 1 Reviews of the effect of interventions (with an example from the diabetes literature)
- Checklist 2 Studies assessing the effect of interventions (with an example from the diabetes literature)
- Checklist 3 Interrupted time series studies assessing the effect of interventions
- Checklist 4 Studies assessing risk factors
- Checklist 5 Studies assessing diagnostic accuracy

2.1 Components of study checklists

Each **study checklist** contains three sections requiring completion by the reviewer:

I. **Descriptive information about the study** covering authors and year of publication, a description of the study intervention, outcomes both beneficial and harmful, other factors that might affect the outcome, characteristics of the study population and setting and the number of groups or sites in the study.

II. **Evaluation criteria for the study** containing the main components of study quality to be considered. The evaluation of the quality of a study, review or guideline, provides information to assist in deciding whether researchers or guideline developers have taken the necessary steps to prevent the over or underestimation of the true effect of interventions, risk factors, diagnostic test accuracy and guideline recommendations.

In this section of the checklists for reviews and studies (Checklists 1-5), criteria are set out in question form to assist the reviewer in identifying whether the researchers have addressed possible opportunities for bias in the design and conduct of the study or review.

In the checklist for intervention guidelines and recommendations (Checklist 6), evaluation criteria are also in the form of questions, which aid the reviewer in assessing whether appropriate procedures have been followed to ensure the guidelines or recommendations will benefit (and not harm) the target population.

Table 2 sets out the codes to be used for the evaluation criteria. The codes are descriptive aids and are not a quantitative scoring system.

Spaces are available on the checklist for reviewers to include comments about how and why they decided on a particular code for each evaluation criterion.

Table 2: Coding for evaluation criteria

Evaluation criteria are coded according to the extent to which the criteria are fulfilled	Code
Criterion entirely fulfilled	a
Criterion mostly fulfilled	b1
Criterion mostly not fulfilled	b2
Criterion not at all fulfilled	c
Criterion not described adequately to classify as a, b1, b2 or c	?
Criterion not applicable	n/a

III. **Overall assessment of the study** allows the reviewer to assess and code the overall quality of the study using the codes in Table 3. Section 2.2 gives advice on using the evaluation criteria to make an overall assessment of quality. Study quality is coded as A, B1, B2, C - these codes are intended to be compatible with those of the Cochrane Collaboration.[6]

Table 3: Codes for overall assessment of quality of study checklists

Low risk of bias	**A**	All or most evaluation criteria from the checklist are fulfilled. Where evaluation criteria are not fulfilled, the conclusions of the study or review are thought very unlikely to alter.
Low - moderate risk of bias	**B1**	Some evaluation criteria from the checklist are fulfilled. Where evaluation criteria are not fulfilled or are not adequately described, the conclusions of the study or review are thought unlikely to alter.
Moderate - high risk of bias	**B2**	Some evaluation criteria from the checklist are fulfilled. Where evaluation criteria are not fulfilled or are not adequately described, the conclusions of the study or review are thought likely to alter.
High risk of bias	**C**	Few or no evaluation criteria fulfilled. Where evaluation criteria are not fulfilled or are not adequately described, the conclusions of the study or review are thought very likely to alter.

Space is available for the reviewer's comments about how and why they decided upon a particular code for the overall assessment of quality. Comments set down explicitly the reasons for a reviewer's assessment of quality. Comments are encouraged as they are useful for explaining reviewers' reasons for coding evaluation criteria, for reassessing studies at a later time, for resolving reasons for differences between reviewers, for additional information about study applicability and areas for further research.

The reviewer should also consider and make a judgement about whether the intervention was responsible for the overall effect shown in the review or study. Comments on future areas for research, the external validity and generalisability of the study including its representativeness, applicability to other settings and relationship to the current organisation of the health system are encouraged. These will assist if the study or review is part of a process to evaluate the effect of the intervention for broader implementation.[12]

2.2 Advice on using the evaluation criteria to make an overall assessment of quality

The **overall assessment of quality** is determined by the separate evaluation criteria and a *judgement* about the relative importance of each source of bias and the extent to which potential biases may collectively influence results.

MERGE is developmental. The exact relationship between the evaluation criteria and the overall assessment of quality is unknown. However we have set out an explicit approach for using the evaluation criteria to make an overall assessment of quality based on sound epidemiological principles.

Examples of how judgements can be made in relation to specific circumstances include:

- A reviewer may not be concerned about lack of blinding for outcomes which are objectively measured, such as all-cause mortality.
- It may not be possible to blind patients to treatment group or health professionals/carers to treatment group. This should be stated in the comments section. The reviewer will need to decide whether the absence of blinding in these cases would bias the result.
- A reviewer may be concerned about a very small non-response rate where the outcome is rare, or where other information in the article suggests that it may be important. These concerns should also be written in the comments section.
- Loss to follow-up or non-response rate is recorded as the percentage of the study population which is not included in the final analysis. There is little research to guide the decision about a sufficiently high response rate to avoid bias. The appraisal team and content experts will need to decide on a cut-off percentage. One possibility is to use 80% response (20% loss) as a cutoff, the rule of thumb figure used by many epidemiologists. Whatever, the cut-off chosen, it needs to be specified in any reports.

Where evaluation criteria are not applicable (code n/a), the overall assessment of quality will be based on the remaining applicable evaluation criteria. The group of reviewers may need to decide explicitly if a particular evaluation criterion is applicable or not for their content area.

If more than 20% of evaluation criteria are coded as "?" or if evaluation criteria judged to be particularly important are coded "?" then the group of reviewers will need to decide explicitly how to proceed. *You* may decide to:

- Assume that if the evaluation criteria are not addressed in the article, then the criteria have not been met.
- Contact the authors of the study for further information so that an assessment of the evaluation criteria can be made.
- If possible, combine (in a meta-analysis) the results of studies and test to see whether including or excluding this study changes the recommendations of a review (sensitivity analysis).
- If the recommendations of a review would change after sensitivity analysis or if the results of the study would be particularly important in determining recommendations, then contact the authors of the study for further information so that an assessment of the evaluation criteria can be made.

2.3 Checklists for reviewing the quality of studies

- Checklist 1 Reviews of the effect of interventions (with an example from the diabetes literature).
- Checklist 2 Studies assessing the effect of interventions (with an example from the diabetes literature).
- Checklist 3 Interrupted time series studies assessing the effect of interventions.
- Checklist 4 Studies assessing risk factors.
- Checklist 5 Studies assessing diagnostic accuracy.

For explanatory notes and definitions to be used in conjunction with the checklists, see Section 5.1.

Checklist 1- Reviews of the effect of interventions

This checklist should be used for assessing the quality of articles presented as reviews, overviews, systematic reviews, or meta-analyses.

DESCRIPTIVE INFORMATION ABOUT THE STUDY	NOTES	DESCRIPTION
Review Identification	Include author, title, reference, year of review and year of publication of review (if available).	
Type of review	Systematic (including meta-analysis) or nonsystematic (maybe described as overviews or summaries).	
What types of study are included in the review?	Randomised Controlled Trials (RCT), Non-Randomised Control Studies (N-RCS), Cohorts, Before and After Studies (BAS), Case Control Studies (C-CS).	
What interventions are considered?		
Are the interventions aimed at individuals or populations?	e.g. drug trial (for individuals), mass media campaign (for populations)	
What outcomes are considered?	i.e. benefits and harms.	
What factors other than the intervention could affect the outcome?	Include potential confounding factors especially for systematic reviews including non-RCTs.	
List the main characteristics of the populations and study settings of the primary studies.	Population characteristics e.g. age, sex, disease characteristics of populations, disease prevalence. Settings e.g. rural, urban, hospital inpatient or outpatient, general practice, community.	

EVALUATION CRITERIA FOR THE REVIEW	COMMENTS	CODE OPTIONS a, b1, b2, c, ?, n/a
Is the search procedure sufficiently rigorous to identify all relevant studies?		
Does the review include all the potential benefits and harms of the intervention?		
Does the review only include randomised controlled trials?		
Apart from study type, is study quality assessed and taken into account?		
Are the data summarised to give a point estimate of effect and confidence intervals?		
Is there an examination of which study population characteristics (disease subtypes, age/sex groups) determine the magnitude of effect of the intervention?		
Is unexplained heterogeneity assessed?		
OVERALL ASSESSMENT OF THE REVIEW	**COMMENTS**	**CODE OPTIONS A, B1, B2, C**
How well (code A, B1, B2, C - see table 3) was this review done to minimise bias? If coded as B1, B2 or C, what is the likely direction in which bias might affect the review results?		
Is the overall effect shown in the review due to the intervention?		
If primary studies are not RCTs, explain if there is any practical or ethical reason why a RCT cannot be done.		
Include other comments concerning areas for further research, applicability of evidence to target population, importance of study to policy development.		

Example using checklist 1 - Reviews of the effect of interventions

This checklist should be used for assessing the quality of articles presented as reviews, overviews, systematic reviews, or meta-analyses.

DESCRIPTIVE INFORMATION ABOUT THE STUDY	NOTES	DESCRIPTION
Review Identification	Include author, title, reference, year of review and year of publication of review (if available).	**Wang P, Lau J, Chalmers T. Meta-analysis of the effects of intensive glycaemic control on late complications of type 1 diabetes mellitus. Online J Curr Clin Trials 1993 May 21;1993 (Doc No. 60).**
Type of review	Systematic (including meta-analysis) or nonsystematic (maybe described as overviews or summaries).	Systematic.
What types of study are included in the review?	Randomised Controlled Trials (RCT), Non-Randomised Control Studies (N-RCS), Cohorts, Before and After Studies (BAS), Case Control Studies (C-CS).	Review of 16 major randomised controlled trials.
What interventions are considered?		Comparison of the effect of intensive and conventional therapy on microvascular complications.
Are the interventions aimed at individuals or populations?	e.g. drug trial (for individuals), mass media campaign (for populations)	Intervention aimed at individuals.
What outcomes are considered?	i.e. benefits and harms.	Benefits: Prevent the occurrence and/or progression of retinopathy and nephropathy. Harms: hypoglycaemia, ketoacidosis.
What factors other than the intervention could affect the outcome?	Include potential confounding factors especially for systematic reviews including non-RCTs.	Considered blood pressure.
List the main characteristics of the populations and study settings of the primary studies.	Population characteristics eg age, sex, disease characteristics of populations, disease prevalence. Settings eg rural, urban, hospital inpatient or outpatient, general practice, community.	529 patients with IDDM treated with continuous subcutaneous insulin infusion or multiple injections, followed up from 8-60 months by specialist diabetes services.

EVALUATION CRITERIA FOR THE REVIEW	COMMENTS	CODE OPTIONS a, b1, b2, c, ?, n/a
Is the search procedure sufficiently rigorous to identify all relevant studies?	Yes	a
Does the review include all the potential benefits and harms of the intervention?	The review deals with benefits for retinopathy and nephropathy but did not include effects on neuropathy. There were problems with evaluating neuropathy because there were no standard definitions or measurement of neuropathy across primary studies. Macrovascular complications and weight gain were not considered.	b1
Does the review only include randomised controlled trials?	Yes	a
Apart from study type, is study quality assessed and taken into account?	Yes - studies not meeting specified criteria were excluded.	a
Are the data summarised to give a point estimate of effect and confidence intervals?	Yes	a
Is there an examination of which study population characteristics (disease subtypes, age/sex groups) determine the magnitude of effect of the intervention?	Some examination of potential 'confounders' e.g blood pressure. Probably not so relevant.	b1
Is unexplained heterogeneity assessed?	Heteogeneity assessed through chi-square test. Significant heterogeneity in trend towards progression of retinopathy after 6 to 12 months of intensive therapy.	a
OVERALL ASSESSMENT OF THE REVIEW	**COMMENTS**	**CODE OPTIONS A, B1, B2, C**
How well (code A, B1, B2, C - see table 3) was this review done to minimise bias? If coded as B1, B2 or C, what is the likely direction in which bias might affect the review results?	Meta-analysis of RCTs. All or most evaluation criteria from the checklist are fulfilled. Low risk of bias.	A
Is the overall effect shown in the review due to the intervention?	Yes	
If primary studies are not RCTs, explain if there is any practical or ethical reason why a RCT cannot be done.	Not applicable.	
Include other comments concerning areas for further research, applicability of evidence to target population, importance of study to policy development.	A very good example of how to do a systematic review with meta-analysis. Results are very likely to be applicable to people with IDDM in NSW and support policies to promote optimal blood glucose control.	

Checklist 2 - Studies assessing the effect of interventions

This set of criteria should be used for assessing the quality of experimental and quasi-experimental studies assessing interventions including randomised controlled studies and trials (RCT), non-randomised controlled studies (N-RCS), cohort and case-control (C-CS) studies, before and after studies (BAS). This set of criteria is not applicable to interrupted time series (ITS) studies (see ITS checklist). Relevant study types are listed in the boxes below each evaluation criteria.

DESCRIPTIVE INFORMATION ABOUT THE STUDY	NOTES	DESCRIPTION
Study Identification	Include author, title, reference and year of publication (if available) and the study timeframe.	
How is the study type described?	Randomised Controlled Trials (RCT), Non-Randomised Control Trials (N-RCS), Cohorts, Before and After Studies (BAS) with/without controls, Case Control Studies (C-CS) - define whether population or hospital based case control study.	
What interventions are considered and how are they implemented?		
Is the intervention aimed at individuals or populations?	e.g. drug trial (for individuals), mass media campaign (for populations)	
What outcomes are considered?	i.e.benefits and harms.	
What factors other than the intervention could affect the outcome?	Include potential confounding factors, differences in baseline characteristics between intervention and control groups.	
What are the characteristics of the population and study setting?	Population characteristics eg age, sex, disease characteristics of the population, disease prevalence. Study Setting eg rural, urban, hospital inpatient or outpatient, general practice, community.	
How many groups/sites in the study?		

EVALUATION CRITERIA FOR THE STUDY					COMMENTS	CODE OPTIONS a, b1, b2, c, ?, n/a
What is the study type?						
RCT	N-RCS	Cohort	BAS	C-CS		
Are study participants well-defined in terms of time, place and person?						
	N-RCS	Cohort	BAS	C-CS		
Is the method of allocation to intervention and control groups/sites independent of the decision to enter the individual or group in the study (adequate allocation concealment)?						
RCT						
What percentage (%) of individuals or clusters refused to participate?					%	
	N-RCS	Cohort	BAS	C-CS		
Are individuals within groups/clusters blind to which intervention group they belong AND are those delivering the intervention (health professionals carers) blind to the intervention group?						
RCT	N-RCS					
Is exposure to interventions measured in a standard, valid and reliable way (avoidance of recall bias)?						
				C-CS		
Is exposure to interventions measured in the same way for both case and control groups? (NB: Objective measures would meet this criteria).						
				C-CS		
Are outcomes measured in a standard, valid and reliable way?						
RCT	N-RCS	Cohort	BAS	C-CS		

EVALUATION CRITERIA FOR THE STUDY					COMMENTS	CODE OPTIONS a, b1, b2, c, ?, n/a
Are outcomes measured in the same way for both intervention and control groups? (NB: Blinding or objective measures would meet this criteria).						
RCT	N-RCS	Cohort	BAS	C-CS		
Are factors other than the intervention e.g. confounding factors, comparable between intervention and control groups and if not comparable, are they adjusted for in the analysis?						
RCT	N-RCS	Cohort	BAS	C-CS		
What percentage (%) of individuals or clusters recruited into the study are not included in the analysis? (Loss to follow up).					%	
RCT	N-RCS	Cohort	BAS	C-CS		
Is the analysis by intention to intervene (treat)?						
RCT	N-RCS		BAS			
Are results homogeneous between sites? (Multicentre/multisite studies only).						
RCT	N-RCS	Cohort	BAS	C-CS		
OVERALL ASSESSMENT OF THE STUDY					**COMMENTS**	**CODE OPTIONS A, B1, B2, C**
How well (code A, B1, B2, C - see Table 3) was the study done to minimise bias? If coded as B1, B2 or C, what is the likely direction in which bias might affect the study results?						
Is the overall effect of the study due to the study intervention?						
If the study type is not an RCT, explain if there is any practical or ethical reason why an RCT cannot be done.						
Include other comments concerning areas for further research, applicability of evidence to target population, importance of study to policy development.						

Example using checklist 2- Studies assessing the effect of interventions

DESCRIPTIVE INFORMATION ABOUT THE STUDY	NOTES	DESCRIPTION
Study Identification	Include author, title, reference and year of publication (if available) and the study timeframe.	**The Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. N.Engl. J.Med. 1993;329:977-86.**
How is the study type described?	Randomised Controlled Trials (RCT), Non-Randomised Control Trials (N-RCS), Cohorts, Before and After Studies (BAS), Case Control Studies (C-CS) - define whether population or hospital based case control study.	RCT
What interventions are considered and how are they implemented?		Intensive therapy aimed at achieving and maintaining glycaemic control as near to normal as possible, while minimising hypoglycaemic episodes. Intensive therapy included flexible insulin doses, frequent monitoring, frequent followup, diet and behavioural change. Conventional therapy aimed for no symptoms of hypoglycaemia or hyperglycaemia, no ketonuria and normal growth and development. No targets for glycaemic control were set.
Is the intervention aimed at individuals or populations?	e.g. drug trial (for individuals), mass media campaign (for populations)	Intervention aimed at individuals.
What outcomes are considered?	i.e. benefits and harms.	Benefits: Prevent the occurence and/or progression of retinopathy, nephropathy, neuropathy. Harms: Hypoglycaemia, ketoacidosis, weight gain.
What factors other than the intervention could affect the outcome?	Include potential confounding factors, differences in baseline characteristics between intervention and control groups.	Baseline characteristics included age, sex, race, duration of IDDM, smoking history, lipids, body weight, blood pressure.
What are the characteristics of the population and study setting?	Population characteristics eg age, sex, disease characteristics of the population, disease prevalence. Study setting eg rural, urban, hospital inpatient or outpatient, general practice, community.	1441 patients aged between 13 and 39 years with IDDM for a mean of 6.5 years. Patients outside specified age range or with hypertension, hypercholesterolaemia, severe diabetic complications or medical conditions were excluded. Setting: inter-disciplinary specialist diabetes services.
How many groups/sites in the study?		Multicentre (29 centres) study across North America.

EVALUATION CRITERIA FOR THE STUDY					COMMENTS	CODE OPTIONS a, b1, b2, c, ?, n/a
What is the study type?					RCT	
RCT	N-RCS	Cohort	BAS	C-CS		
Are study participants well-defined in terms of time, place and person?						n/a
	N-RCS	Cohort	BAS	C-CS		
Is the method of allocation to intervention and control groups/sites independent of the decision to enter the individual or group in the study (adequate allocation concealment)?					Yes	a
RCT						
What percentage (%) of individuals or clusters refused to participate?						n/a
	N-RCS	Cohort	BAS	C-CS		
Are individuals within groups/clusters blind to which intervention group they belong AND are those delivering the intervention (health professionals/carers) blind to the intervention. group?					Not feasible to blind patients or health care providers because of the nature of the intrvention. Maybe not so significant because outcomes were measured blind to intervention group.	b2
RCT	N-RCS					
Is exposure to interventions measured in a standard, valid and reliable way (avoidance of recall bias)?						n/a
				C-CS		
Is exposure to interventions measured in the same way for both case and control groups? (NB: Objective measures would meet this criteria).						n/a
				C-CS		
Are outcomes measured in a standard, valid and reliable way?					Yes	a
RCT	N-RCS	Cohort	BAS	C-CS		

EVALUATION CRITERIA FOR THE STUDY					COMMENTS	CODE OPTIONS a, b1, b2, c, ?, n/a
Are outcomes measured in the same way for both intervention and control groups? (NB: Blinding or objective measures would meet this criteria).					Yes	a
RCT	N-RCS	Cohort	BAS	C-CS		
Are factors other than the intervention e.g. confounding factors, comparable between intervention and control groups and if not comparable, are they adjusted for in the analysis?					Yes	a
RCT	N-RCS	Cohort	BAS	C-CS		
What percentage (%) of individuals or clusters recruited into the study are not included in the analysis? (Loss to follow up).					1% loss to the study with less than 5% loss from all scheduled examinations.	a
RCT	N-RCS	Cohort	BAS	C-CS		
Is the analysis by intention to intervene (treat)?					Yes	a
RCT	N-RCS		BAS			
Are results homogeneous between sites? (Multicentre/multisite studies only).					Yes given that heterogeneity could be explained by the intervention itself varying across different centres e.g. education, nutrition counselling, advice on insulin administration and monitoring, varied with each centre's own program.	a
RCT	N-RCS	Cohort	BAS	C-CS		
OVERALL ASSESSMENT OF THE STUDY					**COMMENTS**	**CODE OPTIONS A, B1, B2, C**
How well (code A, B1, B2, C - see Table 3) was the study done to minimise bias? IF coded as B1, B2 or C, what is the likely direction in which bias might affect the study results.					All or most evaluation criteria from the checklist are fulfilled. Where blinding criteria not fulfilled, it was not feasible to meet these criteria. In this study it was more important to have measurement of outcomes blinded (this criteria was met). This was a major study exploring the issues raised in the guideline with virtually no loss to follow up.	A
Is the overall effect of the study due to the study intervention?					Yes	
If the study type is not an RCT, explain if there is any practical or ethical reason why an RCT cannot be done.					Not applicable	
Include other comments concerning areas for further research, applicability of evidence to target population, importance of study to policy development.					Heterogeneity across centres can be explained through variations in delivery of intensive treatment. It is difficult to explain why the benefits of intensive therapy are seen across the centres while the harms (especially rates of hypoglycaemia) vary enormously between centres. This has implications for generalising the results, especially hypoglycaemia, with intensive treatment to the Australian setting.	

Checklist 3 - Interrupted time series studies assessing the effect of interventions

DESCRIPTIVE INFORMATION ABOUT THE STUDY	NOTES	DESCRIPTION
Study Identification	Include author, title, reference and year of publication (if available) and the study timeframe.	
What interventions are considered and how are they implemented?		
Is the intervention aimed at individuals or populations?		
What outcomes are considered?	i.e. benefits and harms.	
What factors, other than the intervention, could affect the outcome?	Include potential confounding factors, differences in characteristics over time.	
What are the characteristics of the population and study setting?	Population characteristics e.g. age, sex, disease characteristics of the population. Study setting e.g. rural, urban, hospital inpatient, outpatient, general practice, community.	
How many groups/sites in the study?		

EVALUATION CRITERIA FOR THE STUDY	COMMENTS	CODE OPTIONS a, b1, b2, c, ?, n/a
Do any groups act as controls?		
Do factors other than the intervention e.g. confounding factors, remain the same over time? If not, are they adjusted for in the analysis?		
Does the intervention occur at a clearly defined point in time?		
Does the intervention occur independently of other changes in time?		
Are outcomes measured in a standard, valid and reliable way throughout the study?		
Are outcomes measured blind or objectively?		
Could the intervention have affected the data collection?		
What percentage of the data is missing[a] and has it been handled appropriately in the analysis?	%	
Were there sufficient data points to enable reliable statistical inference?[b]		
Was a formal statistical test for change in trend used?		
OVERALL ASSESSMENT OF THE STUDY	**COMMENTS**	**CODE OPTIONS A, B1, B2, C**
How well (code A, B1, B2, C - see table 3) was the study done to minimise bias? If coded as B1,B2 or C, what is the likely direction in which bias might affect the study results?		
Is the overall effect of the study due to the study intervention?		
If the study type is not an RCT, explain if there is any practical or ethical reason why an RCT cannot be done.		
Include other comments concerning areas for further research, applicability of evidence to target population, importance of study to policy development.		

[a] CCEPP recommends 80-100% coverage.
[b] CCEPP recommends 12 data points recorded before the intervention and 12 data points recorded after the intervention.

Checklist 4 - Studies assessing risk factors

This set of criteria should be used for assessing studies aimed at identifying the extent to which characteristics or behaviour of a person, an environmental exposure or the characteristics of a disease alter the risk of an outcome.

DESCRIPTIVE INFORMATION ABOUT THE STUDY	NOTES	DESCRIPTION
Study Identification	Include author, title, reference and year of publication (if available) and the study timeframe.	
What is the study type?	Cohorts, Case Control Studies (C-CS).	
What risk factors are considered?		
What outcomes are considered?	e.g. Disease, surgical rates, death.	
What other factors could affect the outcome(s)?	Include potential confounding factors, demographic characteristics.	
What are the characteristics of the population and study setting?	Personal characteristics e.g. sex, disease characteristics of the population. Study setting e.g. rural, urban, hospital inpatient or outpatient, general practice, community.	

EVALUATION CRITERIA FOR THE STUDY	COMMENTS	CODE OPTIONS a, b1, b2, c, ?, n/a
Are study participants well-defined in terms of time, place and person?		
What percentage (%) of individuals or clusters refused to participate?		
Are outcomes measured in a standard, valid and reliable way?		
Are risk factors and outcomes measured independently (blind) of each other?		
Are all important risk factors included in the analysis?		
What percentage (%) of individuals or clusters recruited into the study are not included in the analysis? (loss to follow up).	%	
OVERALL ASSESSMENT OF THE STUDY	**COMMENTS**	**CODE OPTIONS A, B1, B2, C**
How well (code A, B1, B2, C - see table 3) was the study done to minimise bias? If coded as B1, B2 or C, what is the likely direction in which bias might affect the study results?		
Include other comments concerning areas for further research, applicability of evidence to target population, importance of study to policy development.		

Checklist 5 - Studies assessing diagnostic accuracy

This checklist is concerned with the accuracy of diagnostic tests. It is assumed that the utility of reaching a diagnosis has been adequately established in RCTs.

DESCRIPTIVE INFORMATION ABOUT THE STUDY	NOTES	DESCRIPTION
Study Identification	Include author, title, reference and year of publication (if available) and the study timeframe.	
What is the study type?	Cohort, case-control, cross-sectional analytic studies.	
What is the diagnostic test being considered in this study?		
What is the reference standard being used for comparison? Is there evidence for its validity?		
What condition is the test being used to detect?		
What factors could affect the accuracy of the test?	e.g. special diets prior to some blood tests.	
What are the characteristics of the population and study setting?	Personal characteristics e.g. age, sex, disease characteristics of the population, disease prevalence. Study setting e.g. rural, urban, hospital inpatient or outpatient, general practice, community.	

EVALUATION CRITERIA FOR THE STUDY	COMMENTS	CODE OPTIONS a, b1, b2, c, ?, n/a
Is the test compared with a valid reference standard?		
Are the test and reference standard measured independently (ie blind to the other test result)?		
Is the decision to perform the reference standard independent of the test results (avoidance of verification)?		
If multiple tests are compared, are the tests assessed independently of each other on the same patient or else are they performed on randomly allocated patients?		
Loss to follow up: What % of the described study group are not included in the analysis?	%	
Is the spectrum of patients with and without the disease appropriate for the proposed use of the test?		
OVERALL ASSESSMENT OF THE STUDY	**COMMENTS**	**CODE OPTIONS A, B1, B2, C**
How well (code A, B1, B2, C - see table 3) was the study done to minimise bias? If coded as B1, B2 or C, what is the likely direction in which bias might affect the study results?		
Is this an accurate diagnostic test?		
Include other comments concerning areas for further research, applicability of evidence to target population, importance of study to policy development.		

SECTION 3 SUMMARISING EVIDENCE FROM INDIVIDUAL STUDIES

Once the evidence from the scientific literature has been assessed it should be summarised both quantitatively and qualitatively. The Cochrane Collaboration can give advice on techniques and software for combining results from individual studies (meta-analysis). In this document we present some descriptive approaches for summarising evidence from individual studies. This type of summary can be prepared for use by both consumers and health professionals to promote an understanding of the results from current research and the quality of that research.

3.1 Including or excluding studies from a review of evidence

There is no 'right' or 'wrong' way of doing this. Some options are presented for including studies in a review of evidence:

- Section 1.6 outlines the approach used to locate evidence and incorporate evidence into guidelines for the clinical management of diabetes.

  In step 5 of the method we used a hierarchy of study types starting with systematic reviews of randomised controlled trials and multicentre RCTs to include or exclude the results of an individual study in a review of evidence. This means that studies included in a review of evidence will provide evidence at a particular level (see Table 5).

- After reviewing the evidence available and giving an overall assessment of quality [A,B1,B2,C] for each study, reviewers check how many studies are assessed as A, B1, B2 or C and then decide explicitly which group of studies to include in the review. This means that studies with a high risk of bias will be excluded from the review if better quality studies are available. The decision on cutpoint should be made on the number of studies in each category [A,B1,B2,C] without the studies identified, otherwise there may be bias in choosing the cutpoint to achieve a preconceived result.

- Reviewers decide explicitly that a particular criterion must be met, e.g. response rate must be at least 80% for any study, and then include or exclude a study on that basis.

- Examine whether results vary in different categories of evidence [A,B1,B2,C]. Do the results change when different cutpoints are used?

3.2 Summary of evidence format

Table 4 presents a format for summarising the evidence from the study checklists. This was the format used for summarising the evidence for the guidelines on the clinical management of diabetes. An example of the summary of evidence for the guideline on blood glucose control is given in Section 5.4.

Table 4: Summary of evidence format

TITLE

Of the intervention guidelines and recommendations or purpose of review to which evidence relates.

SUMMARY

These are the major points from the evidence. Main studies are identified. Level of Evidence is rated using Table 5.

SEARCH STRATEGY FOR IDENTIFYING RELEVANT LITERATURE

Search strategies are described including databases, consultation with experts, bibliographies and other references.

QUALITY OF EVIDENCE

Ratings for evaluation criteria and overall assessment of quality of main studies are described; content issues are made explicit. Other important studies are identified.

RESULTS

Benefits and harms from applying the study intervention with quantitative information are presented. Ideally these results should be presented in an easy to read format so that consumers and health professionals can understand the benefits and harms of the intervention. This can be used to assist in decisions to undertake treatment or be involved in an intervention.

CONCLUSION

Major points from the evidence are presented. Comments on generalisability of results, including the representativeness of the study group, its applicability to other settings and the relationship to the organisation of the heatlh system and areas for further research are presented.

REFERENCES

List of individual studies is included in the review of evidence.

SECTION 4 CHECKLIST 6 - INTERVENTION GUIDELINES AND RECOMMENDATIONS

4.1 Uses and components of the guidelines and recommendations checklist (Checklist 6)

The guidelines and recommendations checklist is used for evaluating the validity of the evidence on which intervention guidelines and recommendations are based. This checklist can be used to evaluate the validity of a pre-existing guideline or recommendation or can be used to summarise the validity of guidelines or recommendations during their development. If existing guidelines are evidence-based, it may not be necessary to re-evaluate the original studies. The checklist does not address other evaluation issues such as clinical flexibility, clarity or cost-effectiveness of the guidelines. In this section, the levels of evidence are presented in Table 5 .

The **intervention guidelines and recommendations checklist** contains three sections requiring completion:

I. **A descriptive outline of the guideline or recommendation** covering authors and year of publication, a description of the intervention recommended, guideline users, target group for the intervention and outcomes both beneficial and harmful.

II. **Evaluation criteria for the evidence supporting the validity of the guideline or recommendation** containing the main components to consider in a guideline or recommendation. The same codes are used as in the study checklists (see Table 2). A comments section is available and its use is encouraged.

III. **Overall assessment of the guideline or recommendation** consists of a statement on the level of evidence. Level of evidence is coded as I-V (see Table 5) according to the study types from which the evidence is derived with a secondary code (a,b) for the risk of bias inherent in those studies. Comments can be made about practical or ethical reasons why RCTs cannot be done, applicability of evidence to target populations, areas for further research and importance of the guideline to policy development.

Checklist 6 - Intervention guidelines and recommendations

This checklist should be used for assessing the quality of intervention guidelines and recommendations.

DESCRIPTIVE INFORMATION ABOUT THE GUIDELINE	NOTES	DESCRIPTION
Guideline Identification.	Include summary, title, reference and year of publication.	
What does this guideline recommend?		
Who will use this guideline or recommendation?		
Who is the target group for the interventions recommended?		
What interventions are considered in this guideline or recommendation?		
What outcomes - benefits and harms - are considered?		

EVALUATION CRITERIA FOR THE GUIDELINE	COMMENTS	CODE OPTIONS a, b1, b2, c, ?, n/a
Have all relevent outcomes of the recommendations been considered, covering both potential benefits and potential harms?		
Has an adequate literature review been done to identify, appraise and summarise the evidence from all studies of options relevant to the intervention?		
Are the recommendations consistent with the evidence?		
Do the recommendations explore the trade-off of benefit and harm according to the level of risk for both the outcome and adverse events in different patient subgroups?		
Do the recommendations include the role to be played by individual participant/patient preferences?		
OVERALL ASSESSMENT OF THE GUIDELINE	**COMMENTS**	**CODE OPTIONS LEVEL I-V**
What is the level of evidence for outcomes central to decision making? [Level I-V, see table 5].		
If level of evidence III, IV or V, is there any practical or ethical reasons why randomised trials cannot be done?		
Include other comments concerning areas for further research, applicability of evidence to target population, importance of guideline to policy development.		

Example using Checklist 6 - Intervention guidelines and recommendations

DESCRIPTIVE INFORMATION ABOUT THE GUIDELINE	NOTES	DESCRIPTION
Guideline Identification	Include summary, title, reference and year of publication.	**NSW Health Department Consensus Guidelines for the clinical management of IDDM and NIDDM. NSW Health Department Expert Panel on Diabetes - Guideline 1: blood glucose control - May 1995.**
What does this guideline recommend?		To achieve an HbA_{1c} within 1% of upper limit of normal or achieve blood glucose control as near to this target as possible without producing unacceptable hypoglycaemia. Caution is required in the older population.
Who will use this guideline or recommendation?		Guideline mainly for general practitioners.
Who is the target group for the interventions recommended?		Target group for intervention is non-pregnant adults ($\geq$18 years) with IDDM or NIDDM.
What interventions are considered in this guideline or recommendation?		Assessment of HbA_{1c} every 3-6 months; Assessment of reasons for unsatisfactory control e.g. diet, intercurrent illness, appropriateness of medication, concurrent medication, stress, exercise; Referral to multidisciplinary group of professionals with expertise in the care of people with diabetes who would review, adjust and monitor management and health outcomes accordingly.
What outcomes - benefits and harms - are considered?		Benefits: Prevent the occurrence and/or progression of retinopathy, nephropathy, neuropathy. Harms: hypoglycaemia, ketoacidosis, weight gain.

EVALUATION CRITERIA FOR THE GUIDELINE	COMMENTS	CODE OPTIONS a, b1, b2, c, ?, n/a
Have all relevant outcomes of the recommendations been considered, covering both potential benefits and potential harms?	Benefits from preventing microvascular complications are considered, but benefits from preventing macrovascular complications have not been fully explored in Guideline 1. The major harms have been considered.	b1
Has an adequate literature review been done to identify, appraise and summarise the evidence from all studies of options relevant to the intervention?	Medline search with diverse search strategy, content experts, reference lists and bibliographies.	a
Are the recommendations consistent with the evidence?	Recommendations consistent with the evidence.	a
Do the recommendations explore the trade off of benefit and harm according to the level of risk for both the outcome and adverse events in different patient subgroups?	Partially - risks of hypoglycaemia different in older populations and young children but not fully explored.	b1
Do the recommendations include the role to be played by individual participant/patient preferences?	Yes - covered in Principles of Care and prefacing page of the guidelines.	a
OVERALL ASSESSMENT OF THE GUIDELINE	**COMMENTS**	**CODE OPTIONS LEVEL I-V**
What is the level of evidence for outcomes central to decision making? [Level I-V, see table 5].	Evidence with a low risk of bias was obtained from a systematic review of RCTs where results have been meta-analysed and from a large multicentre RCT.	Ia
If level of evidence III, IV or V, is there any practical or ethical reasons why randomised trials cannot be done?	Not applicable.	
Include other comments concerning areas for further research, applicability of evidence to target population, importance of guideline to policy development.	Evidence related to IDDM only. Prospective studies are underway concerning the prevention of microvascular and macrovascular complications with 'near-normal' blood glucose control in NIDDM. Further research on the hypoglycaemia within the Australian population, including occurrence in subgroups is needed.	

4.2 Level of Evidence

The strength of evidence relates to the quality of studies which support the guideline or recommendation. Study quality indicates how accurately the study is likely to estimate the true effect of an intervention. As discussed in this document, the true effect of a study may be either over or underestimated depending on how well the study was designed and executed. Therefore any classification of the strength of evidence should be based on the potential for bias in the study design and execution.

The well-designed and executed large randomised controlled trial is less likely to be affected by bias than other study designs and therefore provides strong evidence to support an intervention. Well-designed and executed studies using cluster randomisation will provide similarly strong evidence. Where there are a number of similar quality randomised controlled trials, as in a multicentre RCT or where a systematic review of RCTs has been performed, these would provide even stronger evidence. A systematic review should use an adequate literature searching procedure (to ensure all relevant information is identified) and consider the dimensions of quality of the primary studies on which the systematic review is based. Evidence supported by well-designed and executed non-randomised trials, cohort studies and case-control studies can also be strong but may be affected by the limitations of the study design in comparison with randomised controlled trials. Other types of studies may provide less strong evidence to support a guideline or intervention.

The authors acknowledge that the level of evidence needed to support a population-based intervention may not be as high as that needed to support an individual based intervention. This is due to the difficulties in incorporating randomisation and blinding into population studies and the different nature of population interventions. Also, in population-based interventions, the risk to the individual is usually substantially less. In some cases some study types may provide a better level of evidence for a population-based intervention than an individual-based intervention. For example, well-designed and executed time series studies may provide a similar level of evidence for population based interventions as a cohort or case-control study.

There is currently no international consensus on the most appropriate way of classifying the strength of evidence on which guidelines and recommendations are based. QCHOC are using a rating system adapted from the US Preventive Services Task Force (1989) which contains four levels of evidence based on study type with a statement of study quality based on the terms “properly-designed” or “well-designed” (without explicit defining these terms).

To further this debate, we have proposed a way of describing levels of evidence for classifying the quality of studies assessing clinical and public health interventions. Table 5 is based on the US Preventive Services Task Force model[1] and levels of evidence table from the Evaluation Checklist for Evidence-based Guidelines[3] but includes an explicit rating of the risk of bias in the evidence as well as quasi-experimental study designs.

Table 5 Levels of evidence for classifying the quality of studies assessing interventions*

LEVEL OF EVIDENCE	DESCRIPTION OF STUDY TYPES FROM WHICH EVIDENCE IS DERIVED	RISK OF BIAS
I	Systematic review of all relevant randomised controlled trials Large multicentre randomised controlled trials	a Low No unexplained heterogeneity of effect between studies or centres
		b Moderate Unexplained heterogeneity of effect between studies or centres or where heterogeneity is not explored
II	One or more randomised controlled trials and studies**	a Low
		b Moderate
III	Controlled trials without randomisation Cohorts Case-control analytic studies Multiple time series Before and after studies (Preferably from more than one centre or research group)	a Low
		b Moderate***
IV	Other observational studies	
V	Opinions of respected authorities, based on clinical experience, descriptive studies, or reports of expert committees	

* Further research is underway to examine if there is variation in the levels of evidence provided by analytical observation studies and a range of quasi experimental studies.

** If more than one randomised controlled trial or study is available, the results can be combined in a meta-analysis. The combined results would change the level of evidence from II to I.

*** Hospital-based case-control studies would not be rated higher than III b.

Evidence was rated as:

Low risk of bias (code a) if studies from which evidence was derived were well-designed and executed and fulfilled most or all of the criteria from the relevant checklists. Where criteria are not fulfilled, the conclusions of the studies would not alter.

Moderate risk of bias (code b) if studies from which evidence was derived, were properly designed and executed and fulfilled some of the criteria from the relevant checklists. Where criteria are not fulfilled or are not adequately described, the conclusions of the study were unlikely to alter significantly.

SECTION 5 SUPPLEMENTARY NOTES

5.1 Explanatory notes and definitions to be used in conjunction with Checklists:

Definitions in italics are quoted from Last J (ed). A Dictionary of Epidemiology (3rd edition) Oxford University Press. Oxford 1995.

Adequate literature searching procedure/adequate literature review: should include searching electronic databases such as MEDLINE with a variety of search strategies, obtaining articles from content experts and locating further studies from the reference lists of both these sources.

Adequate allocation concealment:[13] The Cochrane Collaboration suggest these criteria for adequate concealment:

- Some form of centralised randomisation scheme, such as having to provide details of an enrolled participant to an office by phone to receive the treatment allocation group.
- Some form of randomisation scheme controlled by a pharmacy.
- Numbered or coded containers, such as in a pharmaceutical trial in which capsules from identical-looking, numbered bottles are administered sequentially to enrolled participants.
- An on-site computer system, given that the allocations are in a locked, unreadable file that can be accessed only after inputting the characteristics of an enrolled participant.
- If assignment envelopes were used, the report should at least specify that they were sequentially numbered, sealed, opaque envelopes.
- Other combinations of described elements of the process that provide assurance of adequate concealment. This may include statements that imply an approach similar to one of those listed above, along with reassuring comments that the person who generated the allocation scheme did not administer it. Some schemes may be innovative and not fit any of the approaches listed above, but still seem to provide adequate concealment. This "other" category is necessary, but is likely to include only a small percentage of all trials deemed to have used adequate allocation concealment.

Inadequate concealment consists of:

- Alternation
- Reference to case record numbers, dates of birth, day of the week, or any other such approach
- Any allocation procedure that is entirely transparent before assignment, such as an **open** list of random numbers or assignments.

Analysis by intention to treat: the analysis compares study and control groups based on the original random allocation regardless of whether individuals in either group received the intervention.

Applicability: extent to which the results of a study or review can be applied to a population or patient group different to that in the original study or review.

Audit: a procedure which *establishes the extent to which a condition, process or performance conforms to predetermined standards or criteria.*

Before-and-after study: study carried out before and after the introduction of an intervention where a group is usually the unit of observation. Where groups or individuals in a before-and-after study are allocated to an intervention or control group, then the study is classified as a non-randomised controlled study.

Benefit(s): an outcome of an intervention which is advantageous for an individual or population.

Bias: systematic errors in the design and execution of a study which may lead to an over- or underestimation of the 'true' effect of an intervention.

Blinded study: *a study in which observer(s) and/or subjects are kept ignorant of the group to which the subjects are assigned...or of the population from which the subjects come. When both the observer and subjects are kept ignorant, we refer to a double-blind study. If the statistical analysis is also done in ignorance of the group to which subjects belong, the study is sometimes described as triple-blind. The intent of keeping subjects and/or investigators blinded, i.e. unaware of knowledge that might introduce a bias, is to eliminate the effects of such biases.*

Case: *a person in the population or study group identified as having the particular disease, health disorder or condition under investigation.*

Case-control study: *a study that starts with the identification of persons with the disease (or other outcome variable) of interest and a suitable control group of persons without the disease.* A population-based case control study is where all the cases come from a defined geographic area and time period, and where controls are a random sample from the same study base. A hospital-based case control study is defined as any other case control study not fitting the definition of a population-based case control study. Population-based case control studies provide better evidence than hospital-based case control studies because they allow the whole spectrum of disease in a population to be examined and they avoid bias from factors which lead a person to be selected as a control. Case control studies are used to estimate relative risk. Case-control studies are useful where the study factor (disease) is rare.

Case report: detailed report on one case usually covering the course of a disease and the response to treatment.

Case series: description of several cases of a given disease (usually covering the course of a disease and the response to treatment).

Clarity: 'guidelines must use unambiguous language, define terms precisely, and use logical and easy-to-follow modes of presentation'.[14]

Clinical flexibility: 'guidelines should identify the specifically known or generally expected exceptions to their recommendations and discuss how patient preferences are to be identified and considered'.[14]

Cluster: the unit of observation is a group rather than an individual e.g. school, worksite, health care team, family, housing block.

Cluster randomisation: the unit randomised is a group (or cluster) rather than an individual (see **randomisation**).

Cohort study: *a study in which subsets of a defined population can be identified who are, have been, or in the future may be exposed or not exposed, or exposed in different degrees, to a factor or factors hypothesised to influence the probability of occurrence of a given disease or other outcome. The alternative terms for a cohort study i.e. followup, longitudinal and prospective study, describe an essential feature of the method, which is observation of the population for a sufficient number of person-years to generate reliable incidence or mortality rates in the population subsets. This generally implies study of a large population, study for a prolonged period (years), or both.* A cohort study is usually the only valid way of assessing absolute risk associated with patient or disease characteristics. A cohort study can be constituted from the non-intervention arm of a randomised trial.

Confidence interval: *the computed interval with a given probability e.g. 95%, that the true value of a variable such as a mean, proportion or rate is contained within the interval.*

Confounding factor: *a variable that can cause or prevent the outcome of interest, is not an intermediate variable and is associated with the factor under investigation.*

Consensus statement: statement on policy or practice based on general agreement or majority of agreement within a group.

Cross-sectional (analytic) study: *a study that examines the relationship between diseases (or other health-related characteristics) and other variables of interest as they exist in a defined population at one particular time....Disease prevalence rather than incidence is normally recorded in a cross-sectional study. The temporal sequence of cause and effect cannot...be determined.*

Descriptive study: *a study concerned with and designed only to describe the existing distribution of variables, without regard to causal or other hypotheses. An example is a community health survey used to determine the health status of people in a community.*

Ecological study: *a study in which the units of analysis are populations or groups of people, rather than individuals. An example is the study of association between median income and cancer mortality rates in administrative jurisdictions such as states and counties.*

Effectiveness: *measure of the extent to which a specific intervention, procedure, regimen or service, when deployed in the field, does what it is intended to do for a specified population.*[15]

Evaluation criteria: specific features of a study or guideline/recommendation relating to quality. Coded as a, b1, b2, c, ? or n/a (see table 2).

Experiment: *a study in which the investigator intentionally alters one or more factors under controlled conditions in order to study the effects of so doing.*

Guideline: 'systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances'.[7]

Harm(s): an outcome of an intervention which is disadvantageous for an individual or population. Also referred to as risks.

Health services evaluation: *the integration of epidemiologic, sociological, economic and other analytic sciences in the evaluation of health services. Components of evaluative health services research are distinguished: evaluation of structure (resources, facilities, manpower); evaluation of process (where, by whom, how health care is provided); evaluation of output (amount and nature of health services provided); evaluation of outcome (results i.e. whether persons using health services experience measurable benefits such as improved survival or reduced disability).*

Heterogeneity: occurs when the results of primary studies on which a meta-analysis is based, or results from different centres in a multicentre trial, differ more than expected by chance. It is usually assessed by a Chi-squared test for heterogeneity. **Unexplained heterogeneity** refers to heterogeneity which is not due to variation in study quality or other characteristics of the study population (disease subtype, age/sex profile) and needs to be taken into account in assessing quality.

Homogeneous: implies there is no significant heterogeneity.

Intervention: public health/health promotion policy or program or clinical treatment regimen aimed at improving health, preventing or minimising disease or changing some other health related characteristic or behaviour.

Intervention Guideline and Recommendation Checklist: checklist used to evaluate the validity of the evidence on which intervention guidelines and recommendations are based. Guidelines have been defined as having **validity** *if, when followed, they lead to the health and cost outcomes projected for them. A prospective assessment of validity will consider the substance and quality of the evidence cited, the means used to evaluate the evidence, and the relationship between the evidence and recommendations.*[14]

Meta-analysis: results from several studies are combined and summarised quantitatively. Meta-analysis usually includes results from randomised controlled trials, however other study types can be included.

Multicentre RCT: randomised controlled trial performed in several different settings e.g. in different hospitals over a broad geographic area.

Non-randomised controlled study: a study or clinical trial where the allocation to the intervention or control groups has not been randomised.

Non-systematic reviews: an explicit and systematic approach has not been used to identify evidence relating to a particular topic. An adequate literature searching procedure has not been used and dimensions of study quality of the primary studies have not been considered.

Objective measures of exposure or outcome: measurement follows standardised procedure which is less open to interpretation by potentially biased observers and study participants e.g. presence of cotinine in saliva, mortality rate, caesarian section rate.

Observational study: *nature is allowed to take its course with changes or differences in one characteristic being studied in relation to changes in other characteristics. Analytic methods such as case control and cohort study designs are called observational studies because the investigator is observing without intervention other than to record, classify, count and statistically analyse results.*

Outcomes: *all the possible results that may stem from exposure to a causal factor or from preventive or therapeutic interventions; all identified changes in health status arising as a consequence of the handling of a health problem.*

Overall assessment: an overall rating on quality of a study, guideline or recommendation using the evaluation criteria. Coded as A, B1, B2, C for study checklists (Table 3). Coded as Level I-V for guidelines & recommendations checklist (Table 5).

Precision:
1. *The quality of being sharply defined or stated, e.g. number of significant digits in the measurement, standard error of measurement, standard deviation of a series of replicate determinations of the same quantity.*
2. *In statistics, precision is defined as the inverse of the variance of a measure or estimate.*

Quality of evidence: degree to which bias has been prevented through the design and conduct of research from which evidence is derived.

Randomisation: a procedure is used so that study units have an equal chance of being allocated to an intervention or control group. (See **adequate allocation concealment**).

Randomised controlled trial or study (RCT): *an experiment in which subjects in a population* (or populations) *are randomly allocated into groups, usually called 'study' and 'control' groups, to receive or not to receive an experimental preventive or therapeutic procedure or intervention. The results are assessed by rigorous comparison of rates of disease, death, recovery, or other appropriate outcome in the study and control groups respectively. Randomised controlled trials are generally regarded as the most scientifically rigorous method of hypothesis testing available in epidemiology.*

Recommendation: advised course of action.

Reference standard: the standard against which a diagnostic test that is being studied can be compared.

Reliability/reproducibility of Guidelines: 'Guidelines are reproducible and reliable if given the same evidence and methods for guidelines development, another set of experts produces essentially the same statements and given the same clinical circumstances, the guidelines are interpreted and applied consistently by practitioners (or other appropriate parties)'.[12]

Representativeness: extent to which the population or patient group in a study or review is comparable to other populations or patient groups.

Retrospective study: a study *that is used to test etiologic hypotheses in which inferences about exposure to the putative causal factor(s) are derived from data relating to characteristics of the persons under study or to events or experiences in their past.*
Case control studies are also referred to as retrospective studies.

Risk factor(s): *an aspect of personal behaviour or lifestyle, an environmental exposure, or an inborn or inherited characteristic, which on the basis of epidemiologic evidence is known to be associated with health-related condition(s) considered important to prevent.*

Selection bias: *error due to systematic differences in characteristics between those who are selected for study and those who are not.*

Sensitivity analysis: *a method to determine the robustness of an assessment by examining the extent to which results are affected by changes in methods, values or variables or assumptions.*

Single centre RCT: randomised controlled trial performed in one setting e.g. in one hospital

Strategy: clinical treatment regimen or public health program (including program aimed at preventing disease or some health-related characteristic).

Strength of Association: extent to which the intervention is associated with the outcome(s) of interest.

Study checklist: one of five checklists used to evaluate the quality of research depending on study type or study purpose.

Study group: in a randomised controlled trial, the group which receives an experimental preventive or therapeutic procedure or intervention. More generally, the group participating in a study.

Study quality: an assessment of the degree to which bias has been prevented through the design and conduct of the study.

Study type: includes randomised controlled trial, cohort, non-randomised controlled trial, population-based case-control, hospital-based case-control, cross-sectional analytic, ecological, descriptive. Randomised controlled trials are the study type of highest quality. Cohort studies, non-randomised trials and population based case-control studies are of higher quality than the remaining study types. Guidelines or recommendations based on evidence from studies other than RCTs should clearly state this and indicate randomised trials should be done. Where RCTs cannot be done for ethical or practical reasons, this should be explained as decisions will then have to rely on evidence from observational studies.

Summary of evidence format: standardised format for summarising evidence after applying study checklists.

Systematic review: an explicit and systematic approach has been used to identify evidence relating to a particular topic. A systematic review should use an adequate literature searching procedure and consider the dimensions of quality of the primary studies on which the systematic review is based. Where studies do not meet defined criteria of quality, these studies should be excluded from the review.

Target population: population receiving an intervention or for whom an intervention is planned.

Time series: a series of (outcome) measurements taken over time with some measurements taken before and some after the introduction of an intervention. Usually more informative than **before-after study** because multiple measurements (compared with 2 measurements in before-after study) allow examination of trends that may indicate causes for an association other than the intervention.

Variability: extent to which the results of different studies differ from each other. Variability may occur because of random error or differences in study design, study setting, participants, interventions, exposure(s) or outcome(s) or in the way these are measured.

Verification bias: estimates of diagnostic accuracy are likely to be biased if, for example, all test positives are verified by the reference standard while only a proportion of test negatives are verified. This bias is avoided if all consecutive subjects who have the test are verified by the reference standard, or subjects are sampled by reference standard results before test results are known.

Well-defined study population/base: entry criteria are well described e.g. place and time; all individuals (or random samples) who fulfil the criteria are included in the study.

5.2 References

1. NH&MRC Quality of Care and Health Outcomes Committee. **Guidelines for the development and implementation of clinical practice guidelines.** AGPS, Canberra, October 1995.

2. Moher D, Jadad A, Nichol G, Penman M, Tugwell P, Walsh S. **Assessing the quality of randomised controlled trials: an annotated bibliography of scales and checklists.** Controlled Clinical Trials 1995;16:62-73.

3. Irwig L, Liddle J, Williamson M. **Evaluation checklist for evidence-based guidelines.** NSW Health Department, Sydney, May 1995.

4. The Cochrane Collaboration on Effective Professional Practice. **The Data Collection Checklist.** March 1996.

5. NHS Centre for Reviews and Dissemination. **Undertaking systematic reviews of research on effectiveness: CRD Guidelines for those carrying out or commissioning reviews.** CRD Report 4. January 1996.

6. Oxman A (Editor). **The Cochrane Collaboration Handbook.** Oxford: The Cochrane Collaboration, 1994.

7. Field M, Lohr K (eds). Institute of Medicine. **Clinical practice guidelines: directions for a new program.** National Academy Press, Washington DC, 1990.

8. Grol R. **Implementing guidelines in general practice care.** Quality in Health Care 1992;1:184-191.

9. Grimshaw J, Freemantle N, Wallace S, Russell I et al. **Developing and implementing clinical practice guidelines.** Quality in Health Care 1995;4:55-64.

10. Grimshaw J, Russell I. **Achieving health gain through clinical guidelines. II. Ensuring that guidelines change medical practice.** Quality in Health Care. 1993;2:243-8.

11. Liddle J, Williamson M. **Incorporating evidence into clinical guidelines for the management of diabetes.** NSW Public Health Bulletin 1995;6:145-147.

12. Gyorkos T, Tannenbaum T, Abrahamowicz M et al. **An approach to the development of practice guidelines for community health interventions**. Canadian J Public Health 1994;85:Supplement 1:S1-S13.

13. Quoted from Oxman A (ed). **Preparing and maintaining systematic reviews (Section VI) The Cochrane Collaboration Handbook, Assessment of methodological quality (Chapter 6)** The Cochrane Collaboration, Oxford, 1994.

14. Field M, Lohr K (eds). Institute of Medicine. **Guidelines for clinical practice: from development to use.** National Academy Press, Washington DC, 1992.

15. Cochrane A. **Effectiveness and efficiency: random reflections on health services.** Nuffield Provincial Hospitals Trust, London, 1972.

5.3 Bibliography

The following publications were used by the authors as a basis for developing the checklists to assess the quality of studies and guidelines, and for classifying levels of evidence. These references will provide readers with additional information on assessing the quality of scientific evidence.

Quality of intervention guidelines

- Eddy DM. **A manual for assessing health practices and designing practice policies: the explicit approach.** American College of Physicians 1992.
- Hayward R, Wilson M, Tunis S, Bass E, Guyatt G. **Users' guides to the medical literature. VIII. How to use clinical practice guidelines. A. Are the recommendations valid?** The Evidence-Based Medicine Working Group. *JAMA* 1995;274:570-4
- Hayward R, Wilson M, Tunis S, Bass E, Rubin H, Haynes R. **More informative abstracts of articles describing clinical practice guidelines.** *Ann Intern Med* 1993; 118;731-737
- Jackson R, Barham P, Bills J, Birch T, McLennan L, MacMahon S, Maling T. **Management of raised blood pressure in New Zealand: a discussion document.** *BMJ* 1993;307:107-110

Quality of reviews of the effect of interventions

- Oxman A. **Checklists for review articles.** *BMJ* 1994;309:648-651
- Oxman A, Cook D, Guyatt G. **Users' guides to the medical literature VI. How to use an overview.** *JAMA* 1994;272:1367-71
- Oxman A, Guyatt G. **Validation of an index of the quality of review articles.** *J Clin Epidemiol* 1991;44:1271-8
- The Cochrane Collaboration on Effective Professional Practice. **The Data Collection Checklist.** March 1996.

Quality of randomised controlled trials and studies

- Chalmers I. **Evaluating the effects of care during pregnancy and childbirth (Chapter 2)** In Chalmers I, Enkin M and Keirse M. **Effective care in pregnancy and childbirth.** Oxford University Press 1989.
- Guyatt G, Sackett D, Cook D. **Users' guides to the medical literature II. How to use an article about therapy or prevention A. Are the results of the study valid?** *JAMA* 1993;270:2598-2601
- Guyatt G, Sackett D, Cook D. **Users' guides to the medical literature II. How to use an article about therapy or prevention B. What were the results and will they help me in caring for my patients?** *JAMA* 1994;271:59-63
- Meinert C, Tonascia S. **Clinical Trials: Design, conduct and analysis.** Oxford University Press. New York 1986
- Richardson W, Detsky A. **Users' guides to the medical literature. VII. How to use a clinical decision analysis. A. Are the results of the study valid?** Evidence-Based Medicine Working Group. *JAMA* 1995;273:1292-5
- Schulz K, Chalmers I, Hayes R, Altman D. **Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials.** *JAMA* 1995;273:408-412

- The Standards of Reporting Trials Group. **A proposal for structured reporting of randomized controlled trials.** *JAMA* 1994;272:1926-30
- The Cochrane Collaboration on Effective Professional Practice. **The Data Collection Checklist.** March 1996.
- NHS Centre for Reviews and Dissemination. **Undertaking systematic reviews of research on effectiveness: CRD Guidelines for those carrying out or commissioning reviews.** CDR Report 4. January 1996.

Quality of observational studies and quasi-experimental studies assessing the effect of interventions

- Levine M, Walter S, Lee H, Haines T, Holbrook A, Moyer V. **Users' guides to the medical literature IV. How to use an article about harm.** *JAMA* 1994;271:1615-9
- Elwood, JM. **Causal relationships in Medicine.** New York, Oxford University Press, 1988.
- The Cochrane Collaboration on Effective Professional Practice. **The Data Collection Checklist.** March 1996.
- NHS Centre for Reviews and Dissemination. **Undertaking systematic reviews of research on effectiveness: CRD Guidelines for those carrying out or commissioning reviews.** CDR Report 4. January 1996.
- Cook TD, Campbell DT. **Quasi-experimentation: Design and analysis issues for field settings**. Houghton Mifflin Co Boston. 1979.

Quality of studies assessing risk factors

- Laupacis A, Wells G, Richardson W, Tugwell P. **Users' guides to the medical literature V. How to use an article about prognosis.** *JAMA* 1994;272:234-7

Quality of studies assessing diagnostic tests

- Irwig L, Tosteson A, Gatsonis C, Lau J, Colditz G, Chalmers T, Mosteller F. **Guidelines for meta-analyses evaluating diagnostic tests.** *Ann Intern Med* 1994;120:667-676
- Jaeschke R, Guyatt G, Sackett D. **Users' guides to the medical literature III. How to use an article about a diagnostic test A. Are the results of the study valid?** *JAMA* 1994;271:389-391
- Jaeschke R, Guyatt G, Sackett D. **Users' guides to the medical literature III. How to use an article about a diagnostic test B. What are the results and will they help me in caring for my patients?** *JAMA* 1994;271:703-7
- Irwig L, Macaskill P, Glasziou P, Fahey M. **Meta-analytic methods for diagnostic test accuracy.** *J Clin Epidemiol.* 1995;48:119-30

Applicability

- Glasziou P, Irwig L. **An evidence based approach to individualising treatment** *BMJ* 1995;311:1356-9
- Rothwell P. **Can overall results of clinical trials be applied to all patients?** *Lancet* 1995;345: 1616-9

5.4 Example of summary of evidence format (evidence relating to blood glucose control guideline)

EVIDENCE RELATING TO GUIDELINE 1: BLOOD GLUCOSE CONTROL

1. SUMMARY

- **Improved blood glucose control can reduce the occurrence and progression of retinopathy, nephropathy and neuropathy in people with *insulin dependent diabetes mellitus (IDDM)*.**

- **By improving blood glucose control, people may experience some weight gain and have more episodes of hypoglycaemia. However, increased episodes of hypoglycaemia are not an inevitable side effect of better blood glucose control.**

These statements are supported by evidence from the Diabetes Control and Complications Trial[1] and a meta-analysis[2,3] of 16 randomised controlled trials. The quality of the evidence is rated at **Level Ia**.

The UK Prospective Diabetes Study[4] will provide more information on whether both microvascular and macrovascular complications can also be prevented in people with *non insulin dependent diabetes mellitus (NIDDM)*.

2. SEARCH STRATEGY FOR IDENTIFYING RELEVANT LITERATURE

- MEDLINE 1966-July 1995 was searched for the key words 'diabetes', 'clinical trial', 'randomised', 'meta-analysis', 'complications' and limited by English language articles and articles relating to humans. Approximately 850 abstracts were identified by this strategy. The full articles were retrieved when the abstract methods indicated a randomised prospective trial of therapies directly related to glycaemic control with microvascular complications as outcome factors.

- Content experts on the Guidelines Working Group provided articles, information on current work, unpublished work and relevant Australian research.[5]

- Relevant articles referenced in the above sources were also identified.

- Diabetes Literature Review Service (DIALOGUE) Adis International (from Third Quarter 1994-First Quarter 1995) was also searched for relevant studies.

- Results from studies which intrinsically introduce greater potential for bias than randomised controlled trials (RCTs) e.g. cohort or cross-sectional studies, were not included in the review because RCTs were available.

- In assessing the quality of evidence, where information was missing from reports or articles, authors were not contacted. We assumed that if criteria were not adequately described, the criteria were not adequately fulfilled.

3. QUALITY OF THE EVIDENCE

3.1 The Diabetes Control and Complications Trial[1] was a North American multicentre (29 centres) randomised controlled trial comparing the effects of intensive therapy with conventional therapy on microvascular complications.

Intensive therapy was aimed at achieving and maintaining glycaemic control as near to normal as possible, while minimising hypoglycaemic episodes. Intensive therapy included flexible insulin doses, frequent monitoring (≥ 4 capillary blood glucose tests per day), frequent follow up (at least once per month), diet and behavioural change. Conventional therapy aimed for no symptoms of hypoglycaemia or hyperglycaemia, no ketonuria and normal growth and development. No targets for glycaemic control were set. Conventional therapy allowed up to 2 insulin injections per day, less frequent monitoring (1 capillary blood glucose test or 1 urine glucose test per day), less frequent follow up (every 3 months), diet advice and conventional diabetes education.

The DCCT followed 1441 patients aged between 13 and 39 years with IDDM for a mean of 6.5 years. Patients outside the specified age range or with hypertension, hypercholesterolaemia, severe diabetic complications or medical conditions were excluded. The mean HbA_{1c} for intensive therapy was 7.2%.

3.1.1 Epidemiological issues

The DCCT had appropriate randomisation procedures including allocation concealment, very low loss to follow up, analysis by intention to treat, comparable baseline features for both conventional and intensive therapy groups and considered the effects of intensive therapy in different subgroups. Measurement of outcomes was 'blind'. It was not feasible to 'blind' participants or health professionals to treatment group. The results showed heterogeneity between centres which could be explained through variations in how both intensive and conventional therapy were delivered in each centre. Diabetes education, nutritional counselling, advice on insulin administration and monitoring varied with each centre's own program. The choice of insulin (e.g. human, pork, beef), onset/duration of action (e.g. rapid acting, intermediate, long acting), the delivery system (multiple injections, continuous subcutaneous insulin infusion) were determined by each centre and individual patient preferences.

3.1.2 Content issues

The way intensive therapy was delivered in the DCCT may not be directly comparable to the Australian setting. Participants in the DCCT were supervised much more closely than people with IDDM in normal clinical practice in North America. Even with this degree of supervision, only 44% of the intensive therapy group achieved near normal glycaemic control while less than 5% maintained this for the life of the trial.

The results presented for the trial relate to the study period, not to the longer term. The effects of near normal glycaemic control on microvascular complications in the long term are not known.

3.2 Meta-analysis of 16 Randomised Controlled Trials by *Wang, Lau and Chalmers*[2,3] reviewed the major randomised controlled trials **prior to the DCCT** which compared the effects of intensive and conventional therapy on microvascular complications. A list of the trials included in the meta-analysis the references for the summary of evidence for this guideline. The combined results related to 529 patients with IDDM. The mean HbA_{1c} for intensive therapy was 7%-10.5% across trials.

3.2.1 Epidemiological issues:

The meta-analysis had a rigorous search procedure to identify all relevant studies. Studies not meeting specified quality criteria were excluded. Only randomised trials were meta-analysed. Data were summarised to give a point estimate of effect and confidence intervals. There was no significant heterogeneity in the effect of intensive therapy on retinopathy (after 2-5 years) or nephropathy.

3.2.2 Content issues:

The meta-analysis concentrated on retinopathy and nephropathy as end points. It did not evaluate neuropathy because there were no standard definitions or measurement of neuropathy across the primary studies.

3.3 Other studies

3.3.1 Cohort studies[6-10] were available which considered risks of developing microvascular complications with improved glycaemic control in adult and paediatric diabetic populations. They provided additional evidence consistent with the RCTs presented.

3.3.2 The *UK Prospective Diabetes Study (UKPDS)*[4] is a prospective, randomised trial in progress which is investigating whether improved blood glucose control will also prevent and/or limit the progression of both microvascular and macrovascular complications in patients with NIDDM. Given that most people in Australia with diabetes have NIDDM, the results of this study will be particularly important in determining management goals.

4. RESULTS

4.1 The Benefits

The DCCT demonstrated that ***intensive therapy*** reduced the occurrence and progression of retinopathy, nephropathy and neuropathy. Participants in the DCCT were stratified into 'primary prevention' and 'secondary intervention' groups at baseline. The results from the DCCT are presented in the following tables for each of these groups.

For the primary prevention group, (IDDM, no retinopathy, urinary albumin excretion < 40 mg/24 hours and duration of diabetes 1-5 years at the start of the trial) the benefits of intensive therapy are reduced risk of complications as indicated below:

Complications of diabetes	Number of cases per year if 1000 people with IDDM are treated with:		Number of cases prevented per year by intensive therapy	Reduction in risk (based on relative risk) of complication by intensive therapy (95% Confidence Interval)
	Conventional Therapy	Intensive Therapy		
≥ 3-step sustained Retinopathy	47	12	35	76% (62-85)
Albumin excretion ≥ 40mg/24 hours	34	22	12	34% (2-56)
Albumin excretion ≥ 300mg/24 hours	3	2	1	44% (0-86)*
Clinical neuropathy**	98	31	67	69% (24-87)

* Not statistically significant.

** Clinical neuropathy was defined as abnormal neurological examination consistent with peripheral sensorimotor neuropathy plus either abnormal nerve conduction in at least 2 peripheral nerves or unequivocally abnormal autonomic nerve testing after 5 years.

For the secondary intervention group (IDDM, very mild-moderate nonproliferative retinopathy, urinary albumin excretion < 200 mg/24 hours and duration of diabetes 1-15 years at the start of the trial) the benefits of intensive therapy are reduced risk of complications as indicated below:

Complications of diabetes	Number of cases per year if 1000 people with IDDM are treated with:		Number of cases prevented per year by intensive therapy	Reduction in risk (based on relative risk) of complication by intensive therapy (95% Confidence Interval)
	Conventional Therapy	Intensive Therapy		
≥ 3-step sustained Retinopathy	78	37	41	54% (39-66)
Macular oedema	30	20	10	23% (0-48)*
Severe nonproliferative or proliferative retinopathy	24	11	13	47% (15-67)
First episode of laser therapy for macular oedema or proliferative retinopathy	23	9	14	51% (21-70)
Albumin excretion ≥ 40mg/24 hours	57	36	21	43% (21-58)
Albumin excretion ≥ 300mg/24 hours	14	6	8	56% (18-76)
Clinical neuropathy**	161	70	91	57% (29-73)

* Not statistically significant.

** Clinical neuropathy was defined as abnormal neurological examination consistent with peripheral sensorimotor neuropathy plus either abnormal nerve conduction in at least 2 peripheral nerves or unequivocally abnormal autonomic nerve testing after 5 years.

The meta-analysis by *Wang et al* of RCTs also demonstrated that ***intensive therapy*** reduced the occurrence and progression of retinopathy and nephropathy. The results from the meta-analysis are presented below:

Complications of diabetes	**Reduction in risk (based on odds ratio) of complication by intensive therapy (95% Confidence interval)**
Progression of retinopathy after 2-5 years of intensive therapy*	51% (15-72)
Progression to proliferative retinopathy or changes requiring LASER therapy after 2-5 years of intensive therapy	56% (13-78)
Nephropathy**	68% (45-81)

* Trend towards progression of retinopathy after 6-12 months of intensive therapy [Odds Ratio: 2.1 95%CI:0.54-8.31;p=0.29 with significant heterogeneity (p=0.046)] which was reversed by 2-5 years of intensive therapy.

** Progression was defined as *an increment in urinary albumin excretion or an increase in glomerular filtration rate if no data on urinary albumin excretion were available*. Changes in blood pressure can affect renal function. Blood pressure (systolic and diastolic) did not change significantly over time in either intensive or conventional therapy groups. The difference between the two groups at the end of the trials was also not significant.

4.2 The Harms

The table below summarises the results from the DCCT, the Wang meta-analysis and Australian studies demonstrating that ***intensive therapy*** increased the number of hypoglycaemic episodes, ketoacidosis* and increased weight in some people.

Potential harms of intensive therapy	**Study**	**Results for intensively treated group**
Hypoglycaemia requiring assistance from a second person	DCCT[1]	43 extra episodes per 100 patient years.
	Wang et al[2,3]	9.1 (95%CI:-1.4,19.6) extra episodes per 100 patient years.
	Australian data[5]	27.2 episodes per 100 patient years (compared with 62 episodes per 100 patient years in DCCT).
Ketoacidosis	DCCT[1]	0.2 extra episodes per 100 patient years.
	Wang et al[2,3]*	12.6 (95%CI:8.7,16.5) extra episodes per 100 patient years.
Weight gain**	DCCT[1]	3.4 extra cases of being 'overweight' per 100 patient years.

* Based on 4 studies using continuous subcutaneous insulin infusion. Continuous subcutaneous insulin infusion was associated with a higher prevalence of ketoacidosis than multiple single (basal bolus) insulin injections. It is usual practice in Australia to use multiple single insulin injections.

** Defined as more than 120% above the ideal weight.

4.3 Other Comments on Hypoglycaemia

Based on DCCT, intensive therapy produced 43 extra episodes of hypoglycaemia per 100 patient years compared with conventional therapy. However, the range of hypoglycaemic episodes in the intensively treated group varied enormously across centres from 0-150 episodes per 100 patient years.[11] For people treated conservatively, the range of hypoglycaemic episodes also varied considerably across centres from 0 - 50 episodes per 100 patient years.[11]

Out of the 29 centres participating in the trial, 16 centres had rates for hypoglycaemic episodes in the intensive group less than the highest rates in the conventional group.[11] Therefore, hypoglycaemia was not an inevitable consequence of better glycaemic control.

Based on meta-analysis by *Wang et al* of RCTs, the rates for hypoglycaemic episodes for different centres also showed enormous variation from 14.7 per 100 patient years ***fewer*** *episodes in the intensively treated group* (95% CI:42.1 per 100 patient years fewer episodes to 12.7 per 100 patient years extra episodes) to 74 per 100 patient years ***extra*** *episodes in the intensively treated group* (95% CI:52.4 per 100 patient years extra to 95.6 per 100 patient years extra episodes in the intensively treated group).

Based on Australian data. A cross-sectional analytic survey of people identified from an Area diabetes register showed lower estimates than the DCCT of hypoglycaemia for intensively treated patients (27.2 episodes per 100 patient years compared with 62 episodes per 100 patient years in the DCCT). Participants were aged 13-39 years, had been diagnosed with IDDM for more than 12 months and were tested for HbA_{1c} within the previous 6 months. The results suggest that intensive therapy was delivered differently in this Area of Australia compared with the United States. Therefore, extrapolating overall hypoglycaemic rates from the DCCT to the Australian setting is not appropriate.

5. CONCLUSION

The major studies presented provide evidence with a low risk of bias that improved blood glucose control reduces the occurrence and progression of microvascular complications in IDDM.

The results of the *UK Prospective Diabetes Study* are required to answer the same questions in NIDDM.

Hypoglycaemia appears to be the main risk associated with improved blood glucose control, however, episodes of hypoglycaemia varied enormously across centres in both the DCCT and the meta-analysis of RCTs presented.

More information is needed about how 'intensive therapy' is delivered across Australia with rates of hypoglycaemia compared.

6. REFERENCES

1. The Diabetes Control and Complications Trial Research Group. **The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus.** *N Engl J Med* 1993;329:977-86

2. Wang P, Lau J, Chalmers T. **Meta-analysis of the effects of intensive glycaemic control on late complications of type 1 diabetes mellitus.** *Online J Curr Clin Trials* 1993 May 21;1993 (Doc No 60)

3. Wang P, Lau J, Chalmers T. **Meta-analysis of effects of intensive blood-glucose control on late complications of type 1 diabetes.** *Lancet* 1993;341:1306-9

4. UK Prospective Diabetes Study (UKDPS) **VIII. Study Design, progress and performance. UK Prospective Diabetes Study Group.** *Diabetologia* 1991 34:877-890

5. Moses R, Rodgers D, Griffiths R. **Diabetic control and hypoglycaemia in the Illawarra Area of NSW, Australia. A comparison with the DCCT.** *J Qualit Clin Pract* 1995;15:89-97

6. Pirart J. **Diabetes mellitus and its degenerative complications: a prospective study of 4,400 patients observed between 1947 and 1973.** *Diabetes Care* 1978;1:168-88,252-61

7. Goldstein D, Blinder K, Ide C, Wilson R, Wiedmeyer H, Little R et al. **Glycaemic control and development of retinopathy in youth-onset insulin-dependent diabetes mellitus. Results of a 12-year longitudinal study.** *Ophthalmology* 1993;100:1125-1132

8. Klein R, Klein B, Moss S, Davis M, DeMets D. **Glycosylated hemoglobin predicts the incidence and progression of diabetic retinopathy.** *JAMA* 1988;260:286402871

9. Klein R, Klein B, Moss S. **Epidemiology of proliferative diabetic retinopathy.** *Diabetes Care* 1992;15:1875-91

10. D'Antonio J, Ellis D, Doft B et al. **Diabetes complications and glycaemic control: the Pittsburgh prospective insulin-dependent diabetes cohort study status report after 5 yr of IDDM.** *Diabetes Care* 1989;12:694-700

11. DCCT Research Group. **Implementation of treatment protocols in the Diabetes Control and Complications Trial.** *Diabetes Care* 1995;18:361-376

LIST OF TRIALS INCLUDED IN META-ANALYSIS BY WANG ET AL

Eschwege E, Job D, Guyot-Argenton C, Aubry J, Tchobroutsky G. **Delayed progression of diabetic retinopathy by divided insulin administration: a further follow-up.** *Diabetologia* 1979;16:13-15

Holman R, Dornan T, Mayon-White V et al. **Prevention of deterioration of renal and sensory-nerve function by more intensive management of insulin-dependent diabetic patients.** *Lancet* 1983;1:204-7

Deckert T, Lauritzen T, Parving H, Christiansen J, Steno Study Group. **Effect of two years of strict metabolic functions in long term insulin-dependent diabetics.** *Diabet Nephropathy* 1984;3:6-10

The Kroc Collaborative Study Group. **Blood glucose control and the evolution of diabetic retinopathy and albuminuria.** *N Engl J Med* 1984;311:365-72

Lauritzen T, Larsen K, Larsen H, Deckert T, Steno Study Group. **Two-year experience with continuous subcutaneous insulin infusion in relation to retinopathy and neuropathy.** *Diabetes* 1985;34(Suppl 3):74-9

Wiseman M, Saunders A, Keen H, Viberti G. **Effect of blood glucose control on increased glomerular filtration rate and kidney size in insulin-dependent diabetes.** *N Engl J Med* 1985;312:617-21

Beck-Neilsen H, Richelsen B, Mogensen C et al. **Effect of insulin pump treatment for one year on renal function and retinal morphology in patients with IDDM.** *Diabetes Care* 1985;8:585-9

Feldt-Rasmussen B, Mathiesen E, Deckert T. **Effect of two years of strict metabolic control on progression of incipient nephropathy in insulin-dependent diabetes.** *Lancet* 1986;2:1300-4

Olsen T, Richelsen B, Ehlers N, Beck-Nielsen H. **Diabetic retinopathy after 3 years' treatment with continuous subcutaneous insulin infusion (CSII).** *Acta Ophthalmol* 1987;65:185-9

Helve E, Laatikainen L, Merenmies L, Koivisto V. **Continuous insulin infusion therapy and retinopathy in patients with type 1 diabetes.** *Acta Endocrinol* 1987;115:313-9

Christensen C, Christiansen J, Schmitz A et al. **Effect of continuous subcutaneous insulin infusion on kidney function and size in IDDM patients: a 2 year controlled study.** *J Diabet Complications* 1987;1:91-5

Dahl-Jorgensen K, Hanssen K, Kierulf P, Bjoro T, Sandvik L, Aagenæs O. **Reduction of urinary albumin excretion after 4 years of continuous subcutaneous insulin infusion in insulin-dependent diabetes mellitus.** *Acta Endocrinol* 1988;117:19-25

Verrillo A, de Teresa A, Martino C, Verrillo L, di Chiara G. **Long-term correction of hyperglycaemia and progression of retinopathy in insulin-dependent diabetes. A five-year randomized prospective study.** *Diabetes Res* 1988;8:71-6

The Kroc Collaborative Study Group. **Diabetic retinopathy after two years of intensified insulin treatment.** *JAMA* 1988;260:37-41

Brinchmann-Hansen O, Dahl-Jorgensen K, Hanssen K, Sandvik L. **The response of diabetic retinopathy to 41 months of multiple insulin injections, insulin pumps, and conventional insulin therapy.** *Arch Ophthalmol* 1988;106:1242-6

Reichard P, Berglund B, Britz A, Cars I, Nilsson B, Rosenqvist U. **Intensified conventional insulin treatment retards the microvascular complications of insulin-dependent diabetes mellitus (IDDM): the Stockholm Diabetes Intervention Study (SDIS) after 5 years.** *J Intern Med* 1991;230:101-8.