Effect of clinical pathway implementation and patients' characteristics on outcomes of coronary artery bypass graft surgery
El-Baz, Noha

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2009

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
CHAPTER 2 | ARE THE OUTCOMES OF CLINICAL PATHWAYS EVIDENCE-BASED?
A CRITICAL APPRAISAL OF CLINICAL PATHWAY EVALUATION RESEARCH
Are the outcomes of clinical pathways evidence-based?
Chapter 2 | Are the outcomes of clinical pathways evidence-based?
A critical appraisal of clinical pathway evaluation research

ABSTRACT
Aim and objective
To evaluate the validity of study outcomes of published papers that report the effects of clinical pathways (CPs).

Method
Systematic review based on two search strategies, including searching Medline, CINAHL, Embase, Psychinfo and Picarta from 1995 till 2005 and ISI Web of KnowledgeSM. We included randomized controlled or quasi-experimental studies evaluating the efficacy of clinical pathway application. Assessment of the methodological quality of the studies included randomization, power analysis, selection bias, validity of outcome indicators, appropriateness of statistical tests, direct (matching) and indirect (statistical) control for confounders. Outcomes included length of stay, costs, readmission rate and complications. Two reviewers independently assessed the methodological quality of the selected papers and recorded the findings with an evaluation tool developed from a set of items for quality assessment derived from the Cochrane Library and other publications.

Results
The study sample comprised of 115 publications. A total of 91.3% of the studies comprised of retrospective studies and 8.7% were randomized controlled studies. Using a quality-scoring assessment tool, 33% of the papers were classified as of good quality, whereas 67% were classified as of low quality. Of the studies, 10.4% controlled for confounding by matching and 59.1% adopted parametric statistical tests without testing variables on normal distribution. Differences in outcomes were not always statistically tested.

Conclusion
Readers should be cautious when interpreting the results of clinical pathway evaluation studies because of the confounding factors and sources of contamination affecting the evidence-based validity of the outcomes.
2.1 Introduction
Clinical pathways have been developed in health care as multidisciplinary care plans that outline the sequence and timing of actions necessary for achieving expected patient outcomes and organizational goals regarding quality, costs, patient satisfaction and efficiency. The concept of clinical pathways (CPs) refers to specific guidelines for care that describe patient treatment goals and define a sequence and timing of intervention for meeting those goals efficiently. They can also be defined as care plans that detail essential steps in patient care with a view to describing the expected progress of the patient. They are also known as critical pathways, integrated care pathway, critical path, care maps and care paths and they are being embraced in many systems.

In an attempt to evaluate the efficacy of integrated care pathways, Campbell et al. posed the question ‘Are clinical pathways effective in improving patient care?’. They used the results of a comprehensive review performed by the National Health Service in Wales in 1996, which was comprised of approximately 4000 references to integrated care pathways and related topics worldwide. The studies that were found mainly described benefits that were experienced and addressed concerns associated with the use of pathways or practical barriers to implementation. Most of the studies they found were uncontrolled ‘before–after’ studies and no randomized controlled studies were found. The authors came to the conclusion that ‘these reports do not provide reliable evidence and publication bias is highly likely, favoring publications reporting favorable experience’.

Every et al. reported that in cardiovascular medicine, although the studies they evaluated were somewhat under-powered, the overall experience had been promising. CPs applied to patients with a cardiovascular disease showed a tendency towards a decreased treatment variation, improved guideline compliance and reduced costs. However, the evidence of the effectiveness of CPs in cardiovascular medicine cannot be generalized because of the insufficient number of controlled studies. Renholm et al. concluded in a review article that clinical pathways had positive effects on patient-care outcome, although some studies did suggest that the use of CPs had no influence on patient-care outcomes, while by the same token they also stated that there was no evidence at all that they had any negative effect.

Similarly, Van Herck et al. concluded that CPs did have a positive effect on patient outcome, but they did not take methodological weaknesses into consideration, because they analysed most of the manuscripts (55.5%) by means of abstracts. Additionally, they expressed their concerns about ‘publication bias since clinical pathways with no, few, or even negative results hardly ever get published’.
Kim et al. conducted a systematic review which focused on the effectiveness of CPs for total knee and total hip arthroplasty\(^6\). They included 11 papers and identified only one randomized controlled study. They reported a decrease in length of stay (LOS) and in costs with either reduced or unchanged rates of complications and either improvement or no change in patient-reported outcomes. Furthermore, they concluded that, although the data in their review supported the effectiveness of CPs, ‘definitive conclusions cannot be made because of methodological limitations’.

Another systematic review was conducted by Kwan et al. regarding CPs for stroke patients\(^6,7\). They included both randomized and non-randomized studies and found no evidence that CPs provided any significant additional benefit over standard medical care in terms of major clinical outcomes (death or discharge destination). Moreover, they concluded that stroke patients in CP groups were more dependent on discharge, while the effect on LOS and hospitalization costs remained unclear.

Despite these uncertainties, CPs have been widely used in many institutions throughout the United States, the United Kingdom and some parts of Europe for patients undergoing various treatments, surgeries or diagnostic procedures. According to the results of the majority of publications in which clinical pathways were compared with standard care, there is an overall tendency to come to the conclusion that CPs are effective in reducing LOS, costs, complication rates and readmissions. On the other hand, CPs may positively affect patients’ health-related functioning, quality of life or patient satisfaction. However, some critical studies\(^1,4,5,8\) have introduced serious doubts about the amount of the evidence and this criticism stimulated us to undertake this study.

The purpose of this study was to appraise the methodological qualities of clinical pathway (CP) evaluation studies in order to evaluate the validity of the evidence for the efficacy CPs and to suggest improvements in future study design. Methodological quality was defined through a set of parameters related to the design and conduct of the study that determines the internal and external validity of the study\(^9-12\).

### 2.2 Methods

Two search strategies were employed to find published studies on the efficacy of CPs. The first strategy involved computerized database searches using Medline, CINAHL, Embase, Psychinfo and Picarta from 1995 till 2005. Studies were identified by a broad range of keywords: clinical pathways, critical pathway, clinical paths and integrated care pathway, effect, adults, paediatric diseases, patient satisfaction, length of stay, complications, readmission, quality of care, quality of life, costs, longitudinal studies, experimental studies, randomized
controlled studies, cohort studies, double-blind methods, systematic review, evaluation and comparative studies.

The second strategy concerned a snowball sampling method using the databases of ISI Web of Knowledge\textsuperscript{SM}. We identified studies that were cited retrospectively in a publication on the effectiveness of CPs, as well as studies that cited the selected papers in the years following their publication.

### 2.2.1 Inclusion and exclusion of publications

Information from abstracts and titles of the papers that were detected using both search strategies was used to include or exclude manuscripts. Papers were excluded if they were:

- Manuscripts that addressed the noun ‘pathway’, but were not related to studies investigating the efficacy of a CPs in terms of specified guidelines or outlines for care that describe patient-treatment goals and define a sequence and timing of interventions to meet those goals efficiently;
- Manuscripts concerning our definition of a CPs which did not provide empirical quantitative results, such as letters to the editor, brief reports, case studies, qualitative designs, opinions of experts, etc..

After exclusion of inappropriate manuscripts, two of the investigators (NEB and BM) independently assessed the remaining publications in full text against two criteria:

1. The study evaluated the efficacy of a CPs by means of quantitative methods (e.g. a meta-analysis or systematic review).
2. The design of the effect study included a CPs group and a control group.

Papers were included if one or both investigators unequivocally considered the publication as appropriate for analysis. Differences were resolved through discussion with reference to a third reviewer (JPvD), if necessary. Methodological appraisal included papers that satisfied both criteria.

### 2.2.2 Assessment of methodological quality

Two reviewers (NEB and BM) independently assessed the methodological quality of the selected papers and recorded their findings with an evaluation tool or scoring system comprising of a set of items for quality assessment from the Cochrane Library and from other publications on quality assessment of studies\textsuperscript{9–11,13}. The quality score was based on eight items that evaluated the following methodological characteristics of individual studies as presented in Table 1.
The appraisal of the overall methodological quality was based on a weighted score in which the similarity of groups at baseline, randomization and adequate statistical methods were assigned three points; the use of power analysis and control of confounders were assigned two points; and assessment by a medical ethics committee, unbiased outcome measure and eligibility criteria were assigned one point. We calculated overall quality scores for each study by summing up the score-weighted ‘yes’ scores.

We established the cut-off point between ‘high’ and ‘low’ quality studies by following the method developed by Verhagen et al.\textsuperscript{12}. They set the cut-off point at 50\% of the maximum achievable score of 16 points, meaning high-quality studies scored ≥8.0 and low-quality studies scored ≤7.0 points. In addition to this dichotomous scale, we arranged the quality scale scores into the following ordinal categories: invalid studies (scores 0–3), weak to medium quality (scores 4–7), good quality (scores 8–11) and high quality (scores 12–16).

### 2.3 Analysis

Analyses were performed using SPSS version 12.0.1 (SPSS Inc., Chicago, IL, USA) and for all tests \( P < 0.05 \) was considered significant. We calculated 95\% confidence intervals for the differences in proportions \textsuperscript{14}. Chi-square and Fisher’s exact tests were used for associations between categorical variables. Apart from the methodological parameters in Table 1, outcomes such as decrease in LOS, costs, readmission and complications and their statistical significance were observed.

### Table 1. The evaluation items used to assess the methodological quality of the studies investigated

<table>
<thead>
<tr>
<th>Question</th>
<th>Yes/No Insufficient information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Was a power analysis performed?</td>
<td>Yes/No Insufficient information</td>
</tr>
<tr>
<td>Was a method of randomization performed?</td>
<td>Yes/No Insufficient information</td>
</tr>
<tr>
<td>Did a medical ethics committee assess the study protocol?</td>
<td>Yes/No Insufficient information</td>
</tr>
<tr>
<td>Were the groups similar at baseline in regards to the most important prognostic indicators?</td>
<td>Yes/No Insufficient information</td>
</tr>
<tr>
<td>Were the eligibility criteria specified adequately?</td>
<td>Yes/No Insufficient information</td>
</tr>
<tr>
<td>Were the operational definitions of outcomes valid?</td>
<td>Yes/No Insufficient information</td>
</tr>
<tr>
<td>Were the statistical tests appropriate to the data?</td>
<td>Yes/No Insufficient information</td>
</tr>
<tr>
<td>Were potential confounders controlled in the study design (e.g. matching) or statistical analysis (multivariate analysis)?</td>
<td>Yes/No Insufficient information</td>
</tr>
<tr>
<td>Were dropouts or losses in follow-up analysed appropriately?\textsuperscript{26}</td>
<td>Yes/No Insufficient information</td>
</tr>
</tbody>
</table>

\*Because of the fact that only 10 studies reported the occurrence of patient dropout or losses in follow-up, this parameter was not incorporated into the final score.

The appraisal of the overall methodological quality was based on a weighted score in which the similarity of groups at baseline, randomization and adequate statistical methods were assigned three points; the use of power analysis and control of confounders were assigned two points; and assessment by a medical ethics committee, unbiased outcome measure and eligibility criteria were assigned one point. We calculated overall quality scores for each study by summing up the score-weighted ‘yes’ scores.

We established the cut-off point between ‘high’ and ‘low’ quality studies by following the method developed by Verhagen et al.\textsuperscript{12}. They set the cut-off point at 50\% of the maximum achievable score of 16 points, meaning high-quality studies scored ≥8.0 and low-quality studies scored ≤7.0 points. In addition to this dichotomous scale, we arranged the quality scale scores into the following ordinal categories: invalid studies (scores 0–3), weak to medium quality (scores 4–7), good quality (scores 8–11) and high quality (scores 12–16).

### 2.3 Analysis

Analyses were performed using SPSS version 12.0.1 (SPSS Inc., Chicago, IL, USA) and for all tests \( P < 0.05 \) was considered significant. We calculated 95\% confidence intervals for the differences in proportions \textsuperscript{14}. Chi-square and Fisher’s exact tests were used for associations between categorical variables. Apart from the methodological parameters in Table 1, outcomes such as decrease in LOS, costs, readmission and complications and their statistical significance were observed.

The appraisal of the overall methodological quality was based on a weighted score in which the similarity of groups at baseline, randomization and adequate statistical methods were assigned three points; the use of power analysis and control of confounders were assigned two points; and assessment by a medical ethics committee, unbiased outcome measure and eligibility criteria were assigned one point. We calculated overall quality scores for each study by summing up the score-weighted ‘yes’ scores.

We established the cut-off point between ‘high’ and ‘low’ quality studies by following the method developed by Verhagen et al.\textsuperscript{12}. They set the cut-off point at 50\% of the maximum achievable score of 16 points, meaning high-quality studies scored ≥8.0 and low-quality studies scored ≤7.0 points. In addition to this dichotomous scale, we arranged the quality scale scores into the following ordinal categories: invalid studies (scores 0–3), weak to medium quality (scores 4–7), good quality (scores 8–11) and high quality (scores 12–16).

### 2.3 Analysis

Analyses were performed using SPSS version 12.0.1 (SPSS Inc., Chicago, IL, USA) and for all tests \( P < 0.05 \) was considered significant. We calculated 95\% confidence intervals for the differences in proportions \textsuperscript{14}. Chi-square and Fisher’s exact tests were used for associations between categorical variables. Apart from the methodological parameters in Table 1, outcomes such as decrease in LOS, costs, readmission and complications and their statistical significance were observed.

The appraisal of the overall methodological quality was based on a weighted score in which the similarity of groups at baseline, randomization and adequate statistical methods were assigned three points; the use of power analysis and control of confounders were assigned two points; and assessment by a medical ethics committee, unbiased outcome measure and eligibility criteria were assigned one point. We calculated overall quality scores for each study by summing up the score-weighted ‘yes’ scores.

We established the cut-off point between ‘high’ and ‘low’ quality studies by following the method developed by Verhagen et al.\textsuperscript{12}. They set the cut-off point at 50\% of the maximum achievable score of 16 points, meaning high-quality studies scored ≥8.0 and low-quality studies scored ≤7.0 points. In addition to this dichotomous scale, we arranged the quality scale scores into the following ordinal categories: invalid studies (scores 0–3), weak to medium quality (scores 4–7), good quality (scores 8–11) and high quality (scores 12–16).

### 2.3 Analysis

Analyses were performed using SPSS version 12.0.1 (SPSS Inc., Chicago, IL, USA) and for all tests \( P < 0.05 \) was considered significant. We calculated 95\% confidence intervals for the differences in proportions \textsuperscript{14}. Chi-square and Fisher’s exact tests were used for associations between categorical variables. Apart from the methodological parameters in Table 1, outcomes such as decrease in LOS, costs, readmission and complications and their statistical significance were observed.

The appraisal of the overall methodological quality was based on a weighted score in which the similarity of groups at baseline, randomization and adequate statistical methods were assigned three points; the use of power analysis and control of confounders were assigned two points; and assessment by a medical ethics committee, unbiased outcome measure and eligibility criteria were assigned one point. We calculated overall quality scores for each study by summing up the score-weighted ‘yes’ scores.

We established the cut-off point between ‘high’ and ‘low’ quality studies by following the method developed by Verhagen et al.\textsuperscript{12}. They set the cut-off point at 50\% of the maximum achievable score of 16 points, meaning high-quality studies scored ≥8.0 and low-quality studies scored ≤7.0 points. In addition to this dichotomous scale, we arranged the quality scale scores into the following ordinal categories: invalid studies (scores 0–3), weak to medium quality (scores 4–7), good quality (scores 8–11) and high quality (scores 12–16).

### 2.3 Analysis

Analyses were performed using SPSS version 12.0.1 (SPSS Inc., Chicago, IL, USA) and for all tests \( P < 0.05 \) was considered significant. We calculated 95\% confidence intervals for the differences in proportions \textsuperscript{14}. Chi-square and Fisher’s exact tests were used for associations between categorical variables. Apart from the methodological parameters in Table 1, outcomes such as decrease in LOS, costs, readmission and complications and their statistical significance were observed.

The appraisal of the overall methodological quality was based on a weighted score in which the similarity of groups at baseline, randomization and adequate statistical methods were assigned three points; the use of power analysis and control of confounders were assigned two points; and assessment by a medical ethics committee, unbiased outcome measure and eligibility criteria were assigned one point. We calculated overall quality scores for each study by summing up the score-weighted ‘yes’ scores.

We established the cut-off point between ‘high’ and ‘low’ quality studies by following the method developed by Verhagen et al.\textsuperscript{12}. They set the cut-off point at 50\% of the maximum achievable score of 16 points, meaning high-quality studies scored ≥8.0 and low-quality studies scored ≤7.0 points. In addition to this dichotomous scale, we arranged the quality scale scores into the following ordinal categories: invalid studies (scores 0–3), weak to medium quality (scores 4–7), good quality (scores 8–11) and high quality (scores 12–16).

### 2.3 Analysis

Analyses were performed using SPSS version 12.0.1 (SPSS Inc., Chicago, IL, USA) and for all tests \( P < 0.05 \) was considered significant. We calculated 95\% confidence intervals for the differences in proportions \textsuperscript{14}. Chi-square and Fisher’s exact tests were used for associations between categorical variables. Apart from the methodological parameters in Table 1, outcomes such as decrease in LOS, costs, readmission and complications and their statistical significance were observed.
2.4 Results
Of the 556 publications that were analysed, 310 papers (55.7%) were not related to our definition of a CPs at all as they addressed either metabolic, molecular or genetic pathways, letters to the editor and editorials or investigated the effect of a pharmacological therapy, or aspects of surgical techniques. Of the remaining sample of 246 publications, 131 manuscripts were excluded because they either did not meet the inclusion criterion regarding required study design (74 papers contained no evaluation study, 52 papers addressed the definition or phases of development of CPs and five were not in English). A total of 115 studies out of 556 investigated the efficacy of CPs according to our definition of CPs and these were used for the assessment of methodological quality (see Figure 1.).

![Flowchart describing study selection and excluded studies](image)

2.4.1 Disease treatments
In the period between 1995 and 2005, 246 publications addressed relevant aspects of CPs in terms of our definition. In Table 2., these publications are shown sorted by disease or intervention category. The most often studied
category was in the field of cardiovascular surgery and diseases in both the gross sample and study sample (21.5% and 17.4%, respectively). Twenty publications evaluating cardiovascular clinical pathways were included in the study sample\textsuperscript{15–34}. The relative number (%) of publications included in the study sample varied between 5% and 16% in the following domains: (1) respiratory diseases, therapy and thoracic surgery\textsuperscript{28,35–51}; (2) gastrointestinal surgery, endoscopic surgery and diseases\textsuperscript{52–70}; (3) orthopedic surgery and multiple trauma\textsuperscript{71–83}; (4) oncological diseases and surgery\textsuperscript{84–92}; (5) neurological trauma, disorders, diseases and pain management\textsuperscript{93–100}; (6) vascular surgery\textsuperscript{15,19,101–105}; and (7) gynaecological diseases, surgery and maternity care\textsuperscript{106–111}.

Categories of diseases or treatment, which represent less than 5% of the study sample, comprised studies on urological diseases, surgery and procedure\textsuperscript{112–116}; psychological disturbances and mental health\textsuperscript{117,118}; metabolic diseases\textsuperscript{119}; paediatric conditions\textsuperscript{120–122}; burn and skin reconstructive surgery\textsuperscript{123,124}; and head and neck surgery\textsuperscript{125,126}.

We tested whether selection bias affected the sample of manuscripts used for methodological analysis with 95% confidence intervals for differences in proportions. The differences between the proportion of papers addressing the effects of CPs on job satisfaction and papers not in English showed under-representation in the final sample (used in the current study) as compared with the total sample of CPs-related publications. The study sample of papers shows a similar distribution across diseases as compared with the gross sample, indicating a good representation of the population of studies published between 1995 and 2005.

\textbf{2.4.2 Designs}

Of the 246 papers, 131 were excluded because of the fact that they were descriptive studies or review articles and only the remaining 115 publications were included based on the criteria of study design. Ten randomized controlled studies were found and the majority (\(n=105\)) were comprised of studies with a retrospective comparative research design or were cross-sectional retrospective studies that compared the differences in patient outcome during a period before and after implementation of CPs. The following retrospective designs were used:

1. A majority of these studies (\(n=96\)) used a historical control group and were conducted at the same hospital (before–after design).

2. Three studies were conducted with a historical control group from a different hospital and at different time periods.

3. Six studies were conducted using concurrent control and experimental groups either in the same hospital, though using separate wards, or in different hospitals.
Are the outcomes of clinical pathways evidence-based?

Ten studies, labeled as randomized controlled studies, comprised studies that randomly selected hospitals where a pathway was implemented or that assigned patients randomly to either pathway care or standard care.

<table>
<thead>
<tr>
<th>Disease and interventions</th>
<th>Gross sample of publications n (%)</th>
<th>Study sample n (%)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Cardiovascular surgery and diseases</td>
<td>53 (21.5)</td>
<td>20 (7.4)</td>
<td>-12.7 - 4.5</td>
</tr>
<tr>
<td>2. Respiratory diseases, therapy and thoracic surgery</td>
<td>20 (8.1)</td>
<td>18 (15.6)</td>
<td>-14.0 - 0.3</td>
</tr>
<tr>
<td>3. Gastrointestinal surgery, endoscopic surgery and diseases (peritonitis)</td>
<td>24 (9.8)</td>
<td>18 (16.8)</td>
<td>-14.8 - 0.1</td>
</tr>
<tr>
<td>4. Orthopaedic surgery and multiple trauma</td>
<td>23 (8.9)</td>
<td>13 (11.3)</td>
<td>-9.2 - 4.4</td>
</tr>
<tr>
<td>5. Oncological diseases and surgery</td>
<td>14 (5.1)</td>
<td>9 (7.8)</td>
<td>-7.8 - 3.5</td>
</tr>
<tr>
<td>6. Vascular surgery</td>
<td>5 (2.2)</td>
<td>7 (6.1)</td>
<td>-1.7 - 2.0</td>
</tr>
<tr>
<td>7. Neurological trauma, disorders, diseases and pain management</td>
<td>18 (7.3)</td>
<td>8 (6.9)</td>
<td>-5.3 - 6.1</td>
</tr>
<tr>
<td>8. Urological diseases, surgery and catheterization</td>
<td>15 (6.1)</td>
<td>5 (4.3)</td>
<td>-3.1 - 6.5</td>
</tr>
<tr>
<td>9. Gynaecological diseases, surgery and maternity care</td>
<td>6 (2.4)</td>
<td>6 (5.2)</td>
<td>-7.3 - 1.7</td>
</tr>
<tr>
<td>10. Paediatric conditions (emergencies, neonatal care, oral rehydration and tonsillectomy)</td>
<td>4 (1.6)</td>
<td>3 (2.6)</td>
<td>-2.3 - 4.2</td>
</tr>
<tr>
<td>11. Head and neck surgery</td>
<td>3 (1.2)</td>
<td>2 (1.8)</td>
<td>-2.2 - 3.8</td>
</tr>
<tr>
<td>12. Burns, decubitus ulcer reconstructive surgery</td>
<td>3 (1.2)</td>
<td>2 (1.8)</td>
<td>-2.2 - 3.3</td>
</tr>
<tr>
<td>13. Psychological disturbances and mental health</td>
<td>10 (4.1)</td>
<td>2 (1.8)</td>
<td>-3.7 - 1.1</td>
</tr>
<tr>
<td>14. Metabolic diseases</td>
<td>5 (2.0)</td>
<td>1 (0.9)</td>
<td>-3.6 - 1.3</td>
</tr>
<tr>
<td>15. Critical pathway evaluation and job satisfaction</td>
<td>6 (2.4)</td>
<td>0 (0)</td>
<td>-4.3 - 0.5</td>
</tr>
<tr>
<td>16. Bone marrow transplant</td>
<td>1 (0.4)</td>
<td>0 (0)</td>
<td>-1.3 - 0.4</td>
</tr>
<tr>
<td>17. Not in English language</td>
<td>7 (2.8)</td>
<td>0 (0)</td>
<td>-4.8 - 0.8</td>
</tr>
<tr>
<td>18. Not related to a specific disease or medical treatment</td>
<td>27 (11.0)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>246 (100)</td>
<td>115 (100)</td>
<td></td>
</tr>
</tbody>
</table>
These randomized controlled studies followed up patients after discharge for outcomes like quality of life, pain, readmission, mortality and complications. One paper assessed patients at baseline, 3, 6 and 12 months\textsuperscript{117}, and another measured patients 1 week after assignment to CP or conventional care and then at 4, 12 and up to 26 weeks\textsuperscript{96–98}. The remaining papers followed up patients after discharge from the health care facility at a time ranging from 10 days up to 12 weeks\textsuperscript{42,45,56,72,127,128}.

### 2.4.3 Randomization and matching

We detected 12 retrospective studies (10.4\%) that controlled for confounding through matching, of which three studies used a random sample from a CP group which was matched with controls from the pre-pathway period\textsuperscript{25,30,48,50,58,79,81,107,110,111,122,129}. Furthermore, 10 randomized controlled studies were found, of which two studies randomly assigned hospitals either to implement a CP or to remain on standard care\textsuperscript{45,127}. Eight studies randomly assigned patients to either a CP or standard care\textsuperscript{42,56,72,96–98,117,128}.

### 2.4.4 Power analysis

The question of what sample size to use constitutes a crucial part of any research proposal. However, only 16.5\% (\(n=19\)) of the studies conducted a power analysis to determine in advance the required number of observations which would be sufficient to provide the required precision of results. Among the 115 studies, 25\% of the samples were very small (\(n<50\)), 25\% ranged from 51 to 100 patients, 25\% ranged from 100 to 200 patients and 25\% had samples greater than 200 patients in either the CP or the control group. No statistically significant association was found between sample size (\(n<100\) vs. \(n>100\)) and performance for a statistical power analysis (Chi-square, \(P=0.56\)).

### 2.4.5 Control for confounding

In less than half of the studies (42.1\%), the authors did pay attention to the problem of controlling for potential confounders and either applied direct control with randomization or matching (17.5\%), and/or used a control for confounders (e.g. co-morbidity, age and gender) in regression models (24.6\%). The other (57.9\%) did not take the potential risk of confounders into account at all.

Selection bias may occur when inclusion and exclusion criteria are obscure. In 57 of the studies that were appraised (49.6\%), misleading conclusions were prone to arise based on the fact that patients in ‘experimental’ groups differed from the control group patients regarding characteristics such as differences in
Are the outcomes of clinical pathways evidence-based?

age, gender, disease severity or co-morbidity. The other half of the studies meticulously described criteria for inclusion and exclusion.

2.4.6 Accuracy and validity of outcome measures
Length of stay was evaluated in 108 publications (93.3%). However, more than a quarter (28.1%) of these studies gave no accurate or meticulous description of its operationalization or a clear description of the way it was assessed. Costs and hospital charges were assessed in 73 papers (63.5%), among which 62 (53.9%) stated a clear description of the charges and costs calculated.

• *Readmission rates* were calculated in 53 papers (46.1%), of which 50 (43.5%) precisely defined readmission within a time frame.
• *Complications* were evaluated in 70 papers (60.9%), of which 64 (55.6%) defined and clearly stated the complications.
• *Quality of life* was assessed in 10 (8.7%) of the studies with a validated measure. Functional health-related functioning was measured in six (5.2%) studies, and both quality of life and health status were measured in two papers. Three studies (2.6%) assessed psychological distress (anxiety and depression). However, only two studies used a validated measure: the Hospital and Anxiety and Depression Scale.
• *Patient satisfaction* was assessed in 15 studies (13.0%), but was measured with a multi-item tool in 13 studies (11.3%). *Work satisfaction* was evaluated in four (3.5%) of the studies, and three studies presented an accurate description of this construct. Clinical quality-of-care indicators were evaluated in 57 (49.6%) papers and were accurately defined in all cases.

2.4.7 Appropriateness of statistical methods
More than half (59.1%) of the studies adopted parametric statistical tests without question, but the rest (40.9%) tested variables over normal distribution and, depending on the outcome, used non-parametric tests. Reduction of LOS, costs, readmission rates and number of complications belong to the most relevant targets for implementing CPs.

However, decreases in LOS, costs, readmission rates and number of complications were not statistically tested in 12.3%, 28.8%, 20.8% and 27.1% of the studies, respectively. There were studies that used a statistical test to decide whether a difference between a CP group and a control group was due to sample fluctuation, but they also reported other differences without this test. For the main outcome parameters, LOS, readmission rates, costs and number of complications (65%) of the studies tested for each outcome (23%) did not test for
all differences between CPs and controls, while (12%) did not apply any test at all.

2.4.8 Quality of studies related to statistically significant outcomes

We found 92 publications that reported a decrease in LOS and 60 that reported a decrease in costs. All the good-quality studies reported a statistically significant result in both LOS and costs. However, among the low-quality studies (84%) of the papers reported a reduction in LOS that was statistically significant, and only 68% of the publications reported a decrease in costs which was statistically significant (Fisher’s exact test, $P=0.02$ and $P=0.03$, respectively). There was no association between quality of the studies and the statistical significance of the reduction in complications and in readmission rates (see Table 3).

<table>
<thead>
<tr>
<th>Statistically significant reduction in:</th>
<th>Low quality</th>
<th>Good quality</th>
<th>Total</th>
<th>$P^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length of stay</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>52 (83.9%)</td>
<td>30 (100%)</td>
<td>82</td>
<td>0.02</td>
</tr>
<tr>
<td>No</td>
<td>10 (16.1%)</td>
<td>0</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>62</td>
<td>30</td>
<td>92</td>
<td></td>
</tr>
<tr>
<td>Costs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>32 (68.1%)</td>
<td>13 (100%)</td>
<td>45</td>
<td>0.03</td>
</tr>
<tr>
<td>No</td>
<td>15 (31.9%)</td>
<td>0</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>47</td>
<td>13</td>
<td>60</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Comparison of statistically significant results on the two main outcome measures between studies with low and good methodological quality

*Fisher’s exact test

2.4.9 Overall quality related to other study characteristics

According to the dichotomous threshold of Verhagen et al., one-third of the papers (33%) analysed were classified as high-quality papers, while 67% were classified as studies of low quality. However, the ordinal quality scale showed that 35.7% of the studies reflected low quality, that is, invalid studies (31.1%) were appraised as weak- to medium-quality studies, and 24.3% as good-quality studies. Finally, 8.7% of the sample reflected very good quality.

No statistically significant associations were found between the quality of the studies and the sample size (Mann–Whitney U-test/Wilcoxon W-test, $Z=-0.48$, $P=0.63$). Across the diseases shown in Table 1., the differences between proportions of low- or high-quality studies were not statistically significant. Furthermore, the quality of the studies was not associated with treatment in terms of surgery versus non-surgery (Fisher’s exact test, $P=0.31$).
Data extracted from hospital records were used in 81% of the studies and 19% used self-reported questionnaires or interviews in combination with data from the hospital records.

Both dichotomous and ordinal categorization confirmed that studies which qualified as ‘good quality’ were more likely to use patient record information in combination with self-report questionnaires or interviews (Fisher’s exact test, $P=0.01$). Because of the fact that the majority of studies (91.3%) used electronic databases, the question of whether dropouts were analysed appropriately did not play a significant role in our analysis. A significant association was found between the quality of the study and its design. Only 27% of the studies with a retrospective design were classified as of good quality (Fisher’s exact test, $P=0.0001$), whereas all randomized controlled studies were of high quality.

### 2.5 discussion

The majority of the publications on CP we analysed were classified as studies of low quality (67%), which raise questions about the validity of the evidence for the implementation of CP in the last decade.

The methodology of studies assessing the efficacy of CPs has been criticized in regard to their research designs, poor reporting of the methodology and under-powered sample sizes\(^1,3,5–8\). Therefore, we performed an appraisal and analysis of the methodological qualities of a large number of CP evaluation studies in order to investigate the validity of their outcomes.

In relation to the internal and external validity of the studies in the sample, our analysis revealed several factors that influenced the quality of CP evaluation. Most studies in our sample concerned non-randomly selected small patient populations without a power analysis\(^19,23,25,36,52,54,56,58,59,61,76,81,86–88,91,99,107,111,118,122,123,126,128\).

Furthermore, in many studies selection bias has occurred. Patients selected for a CP were likely to differ from patients managed with standard care using characteristics known to be associated with LOS and hospital delay, and consequently included costs such as age and co-morbidity. In half of the study sample (49.6%), the inclusion and exclusion criteria of patients were obscure and difficult to relate to the population’s characteristics. According to our findings, only 12 retrospective studies (10.4%) controlled for selection bias by matching, out of which three studies matched a random sample from a CP group with controls from the pre-pathway period group\(^25,30,48,50,58,79,81,107,110,111,122,129\).

In regard to the outcomes measured, our analysis also revealed that most studies focused on cost issues and reductions in LOS. However, clinically relevant outcomes such as mortality, discharge disposition, quality of care as seen through the eyes of the patient, psychological distress (anxiety), care
dependency and use of health services after discharge were largely ignored. A few studies focused on outcome in terms of: (1) destination of discharge\textsuperscript{21,24,69,72,78,98,101,104,122}; (2) delay\textsuperscript{44,109}; (3) patient education\textsuperscript{32,124}; (4) quality of care\textsuperscript{22,74}; and (5) psychological distress\textsuperscript{48,98}.

In relation to the use of appropriate statistical tests, more than half (59.1%) of the studies adopted parametric statistical tests without question, but 40.9% tested variables on normal distribution and, depending on the outcome, used non-parametric tests. Before–after differences in frequently measured outcomes such as LOS, costs, readmission and complications were statistically tested in few of the studies. Other studies used statistical tests to decide whether differences between groups were related to sample fluctuations, while also reporting other differences without statistical tests. After reviewing all papers included, it was clear that there was an inevitable risk of misrepresentation of the true costs.

Reduced LOS in the hospital may lead to admission to another health care facility, for example, a rehabilitation unit or extended care facility. Although costs related to these services are not considered as a part of hospital charges, they still add to direct health care costs and may also lead to out-of-pocket expenses for the patient. Furthermore, studies that report reduction in both LOS and costs do not take into consideration that both are interrelated, which leads to contamination in statistical analysis.

Moreover, from a methodological point of view both the investigators (who assessed the outcomes) and the health care providers should have been blinded to the use of a CP since this might have biased their observations and assignment procedure.

In the light of these findings, it can be concluded that readers must be extremely cautious when interpreting the results of CP evaluation studies because of the confounding factors and the sources of contamination affecting the internal and external validity of most of the published studies.

\textbf{2.6 Recommendations for future research}

After reviewing a large number of CP evaluation studies and having stated our conclusions, we recommend that:

\textbf{1. More (randomized) controlled studies should be conducted, in which patients are randomly assigned to the condition of either a pathway or standard should be conducted. However, such randomized controlled studies in the same hospital invite contamination, because many of the same doctors, as well as care staff, are involved in treating the same population of patients. To avoid such Hawthorne effects, we suggest establishing multi-centre trials with randomization after pre-stratification of confounding factors (e.g. gender, co-}
Are the outcomes of clinical pathways evidence-based?

morbidity) with a clearly defined method of randomization, concealment of allocation or blinding with the appropriate balancing method 130.

2. Standardization of the total direct costs is specified by clearly defined cost components and a standardized operational definition of LOS. LOS should not be confined to the hospital setting, but should be extended to include whether patients are discharged home or to an extended health care facility, and should include whether this is a permanent or temporary arrangement. An accurate calculation in this case will reflect the true effect of CPs on LOS and subsequent related costs.

3. More attention should be paid to measuring relevant patient outcomes such as quality of life, hospital anxiety, patient expectations and satisfaction with standardized validated tools, which reflect the true effect with use of appropriate statistical methods. Until the results of good-quality research are more widely available, CPs should be considered as a potentially effective (but not evidence-based) practice to improve patient care.
2.7 Reference List

Are the outcomes of clinical pathways evidence-based?


Are the outcomes of clinical pathways evidence-based?


Chapter 2


Are the outcomes of clinical pathways evidence-based?


Are the outcomes of clinical pathways evidence-based?