METHODOLOGICAL CHALLENGES IN RARE DISEASE GUIDELINES

METHODOLOGICAL CHALLENGES IN EVIDENCE GATHERING AND ASSESSMENT FOR GUIDELINE DEVELOPMENT IN RARE DISEASES: AN EXAMPLE FROM HEMOPHILIA

By CINDY HOI TING YEUNG, B.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the

Requirements for the Degree Master of Science

McMaster University

McMaster University MASTER OF SCIENCE (2016) Hamilton, Ontario

TITLE: Methodological challenges in evidence gathering and assessment for guideline development in rare diseases: an example from hemophilia

AUTHOR: Cindy Hoi Ting Yeung, B.Sc.

SUPERVISOR: Dr. Alfonso Iorio

NUMBER OF PAGES: ix, 146146

Abstract

The guideline development process can be challenging when diseases are rare. The development of the NHF-McMaster Guideline on Care Models for Hemophilia Management presented an opportunity to identify methodological challenges for guideline development in rare diseases. Several methodological challenges were identified in the evidence gathering and assessment stages. Eight challenges were presented in the development of a systematic review on care models for hemophilia management. The barriers to conducting the systematic review were mostly due to the paucity of high quality evidence in hemophilia care models. Due to the paucity of high-quality evidence typical of a rare condition such as hemophilia, indirect evidence from other chronic conditions were sought through an overview of reviews. Seven challenges were identified in the development of an overview on integrated multidisciplinary care for the management of chronic conditions in adults. The barriers were mainly due to unestablished methodology for conducting overviews, and the challenge of applying this evidence in the context of hemophilia for the guideline. To overcome the methodological challenges with evidence gathering and assessment for rare disease guideline development, the decision-making process to derive solutions were transparently presented. Overall, the methodological challenges as well as apparent facilitators from a rare disease setting are shown to be related to the barriers and facilitators at the research, clinical, and guideline development phase. As a result, using an example from hemophilia, this thesis has demonstrated that it is possible to develop high quality guidelines for rare diseases.

Acknowledgements

I would first like to thank my thesis supervisor, Dr. Alfonso Iorio throughout my thesis work over the years. I am truly appreciative of his support as a mentor, providing guidance throughout my thesis and key life lessons that I can take outside of research. Dr. Alfonso Iorio has always challenged me to reach my potential and seize opportunities that arise. Using his words, "You will never get a chance of winning the lottery if you don't purchase a ticket." I am very grateful to Dr. Alfonso Iorio's for insisting that I buy each ticket.

I would also like to thank my thesis committee members, Drs. Nancy Heddle and Jason Busse for their feedback and guidance. Their expertise and critical insight helped to guide and refine the writing of this thesis. I also thank Dr. Robby Nieuwlaat for his supportive role as external examiner.

Many thanks go to the experts who were involved in the preliminary work, research process, and completion of this thesis. In particular, the tremendous amount of support and guidance offered by Dr. Menaka Pai and Dr. Nancy Santesso are beyond appreciation. Thank you for always bringing forth a different perspective to my thesis work and sharing your immense knowledge and expertise. I would also like to thank Tamara Navarro-Ruan for her guidance through systematic review methodology, especially at the start of this thesis.

Finally, I must express my very profound gratitude to my parents, sisters, and friends for providing me with their unwavering support and continuous encouragement throughout my years of study. This accomplishment would have not been possible without them.

Table of Contents

Chapter 1. Introduction
1.1 Clinical practice guidelines and rare diseases....1
1.2 Development of clinical practice guidelines for rare diseases: an example from hemophilia....3
1.2.1 Guideline development rationale....4
1.2.1.1 What is hemophilia and why it is considered a complex disease....4
1.2.1.2 Care models in the management of hemophilia....5
1.2.1 Guideline development process and initial stages....6
1.2.2 The guideline development team....8
1.3 Thesis outline....9
1.3.1 Thesis overview and rationale....9
1.3.2 Thesis objective....12

Chapter 2. Developing a systematic review of care models for hemophilia management
2.1 Introduction....13
2.2 Objectives....13
2.3 Methodology, challenges, and solutions....13
2.3.1 Searching for studies....13
2.3.1.1 Designing the search strategy....14
2.3.2 Selection of studies and collecting data....18
2.3.2.1 Results from the search and records for screening....18
2.3.2.2 Developing the screening criteria....21
2.3.2.3 Categorizing studies that focus on care models....22
2.3.2.4 Data extraction and using results from studies with different objectives....24
2.3.3 Assessing quality of evidence in included studies....28
2.4 Results....34
2.5 Conclusion....40

Chapter 3. Developing an overview of reviews regarding integrated multidisciplinary care for the management of chronic conditions in adults
3.1 Introduction....42
3.2 Objectives....43
3.3 Methodology, challenges, and solutions....43
3.3.1 Scope of the overview of reviews....43
3.3.1.1 Objectives of the overview of reviews....44
3.3.1.2 Overall search strategy plan....45
3.3.1.3 Definition and scope of integrated care....51
3.3.2 Searching for reviews....52
3.3.2.1 Clarifying the unit of analysis....53
3.3.2.2 Updating the meta-review....54
3.3.2.3 Structure of the search strategy....54

3.3.3 Selection of reviews and collecting data 54
3.3.3.1 Developing the screening criteria 54
3.3.3.2 Data extraction 57
3.3.4 Assessment of risk of bias of the included reviews 59
3.3.5 Assessment of quality of evidence of each outcome 62
3.3.6 Assessment of quality of evidence for the guideline 67
3.4 Results 68
3.5 Conclusion 77

Chapter 4. Methodological challenges and facilitators in the development of a guideline in rare diseases
4.1 Introduction 80
4.2 Summary of the methodological challenges and facilitators 80
4.2.1 Barriers in clinical care, research, and guideline development 81
4.2.4 Overall facilitators 83
4.3 Conclusion to thesis 85

List of figures

Figure 1. Observed care model organization in the United States. 97
Figure 2. Steps in the guideline development process. 98
Figure 3. Overview of the methodology to retrieve, evaluate, and synthesize evidence for the NHF-McMaster Guideline on Care Models for Hemophilia Management 99
Figure 4. The influence of adding comprehensive care center and hemophilia treatment center terms to the search strategy 100
Figure 5. Classification system of hemophilia models of care to reduce screening inclusion criteria ambiguity 101
Figure 6. PRISMA flow diagram for systematic review on care models in the management of hemophilia 102
Figure 7. PRISMA flow diagram for overview of reviews on integrated multidisciplinary care for the management of chronic conditions in adults 103

List of tables

Table 1. Facilitators of and barriers to clinical care, research, and guideline development in rare diseases 104
Table 2. NHF-McMaster Guideline on Care Models for Hemophilia Management question 1 106
Table 3. NHF-McMaster Guideline on Care Models for Hemophilia Management question 2 107
Table 4. Examples of retrieved studies organized by the developed hemophilia care models classification system and data use for the systematic review. 108
Table 5. Description of included non-randomized comparative studies 110
Table 6. Risk of bias summary by non-randomized comparative study assessed by the ACROBAT-NRSI tool 113
Table 7. GRADE evidence profile for summary of findings from non-randomized comparative studies 114
Table 8. Outcome data from non-randomized non-comparative studies 117
Table 9. Characteristics of the included reviews and studies of the overview of reviews 119
Table 10. GRADE summary of findings table of all reviews by outcome and by chronic condition 123
Table 11. GRADE evidence profile for care models in hemophilia (data from other chronic conditions) 127

Appendices

Appendix A. Care models for hemophilia management search strategies in MEDLINE, EMBASE and CINAHL 130
Appendix B. Description of single-arm non-comparative studies 132

List of abbreviations

6MWT	6-Minute Walking Test
ACROBAT-NRSI	A Cochrane Risk of Bias Assessment Tool: for Non-Randomized Studies of Interventions
AQLQ	Asthma Quality of Life Questionnaire
CENTRAL	Cochrane Central Register of Controlled Trials
CI	Confidence Interval
CINAHL	Cumulative Index to Nursing & Allied Health Literature
COPD	Chronic Obstructive Pulmonary Disease
EQ4D	EuroQol 4-Dimensions Questionnaire
FEV1	Forced Expiratory Volume in 1 second
FVC	Forced Vital Capacity
GRADE	Grading of Recommendations, Assessment, Development and Evaluation
HR	Hazard Ratio
HTA-DM	Health Technology Assessment-Disease Management
HTC	Hemophilia Treatment Center
MASAC	Medical and Scientific Advisory Council
MD	Mean Difference
MeSH	Medical Subject Headings
MLHFQ	Minnesota Living with Heart Failure Questionnaire
OR	Odds Ratio
PICO	Population, Intervention, Comparator, and Outcomes
RCT	Randomized Controlled Trial
ROBINS-I	The Risk of Bias In Non-randomized Studies – of Interventions
ROBIS	Risk Of Bias In Systematic Reviews
RR	Relative Risk
SGRQ	St. George's Respiratory Questionnaire
SMD	Standardized Mean Difference
US	United States
WHF	World Health Federation

Declaration of academic achievement

The contents of this document has been completed by Cindy Hoi Ting Yeung. The thesis describes the methodological challenges and solutions in the development of the NHF-McMaster Guideline on Care Models for Hemophilia Management. The contents focus on two [1, 2] of the six published papers that describe the guideline process in the journal, *Haemophilia*. The first manuscript describes the systematic review on care models on the management of hemophilia [1]. The second manuscript describes the overview of reviews on integrated multidisciplinary care for the management of chronic conditions in adults [2]. For both manuscripts, Cindy Hoi Ting Yeung contributed to performing the literature search, the quality of evidence assessments, and the writing of the paper. Processes beyond the evidence gathering and assessment stages (choosing the guideline scope, question generation, and conflict of interest assessments) were completed by the other guideline development team members. Therefore, the contributions of Dr. Menaka Pai, Dr. Nancy Santesso, Dr. Holger J Schünemann, and Dr. Alfonso Iorio in the preliminary work, research process, and completion of this thesis are recognized and appreciated.

Chapter 1. Introduction

1.1 Clinical practice guidelines and rare diseases

Clinical practice guidelines that are developed according to a rigorous methodological standard which combines scientific evidence, clinician expertise and knowledge, and patient values, have the potential to provide systematic aid to making complex medical decisions for the management of patient populations. Specific patient populations, notably those with rare diseases, often require making complex clinical decisions for the individual patient. However, to make these clinical decision involves having clinical experience and evidence to support clinical care decisions, both of which takes significantly more time to gain for rare conditions than for common diseases. Thus, clinical practice guidelines are potentially very helpful in supporting clinical decision making for rare diseases. Although there is no consensus on a definition for rare diseases, they are characterized by a low prevalence in the population, however, they can cumulatively affect millions of people of all ages globally [3]. There is thought to be over 7,000 rare diseases, with over 25 million American and 30 million European Union citizens affected [4, 5]. The United States (US) Food and Drug Administration Orphan Drug Act considers a disease to be rare if it has a prevalence of fewer than 200,000 affected individuals in the US (1 in 1,500 people), while the European Organ Drug Regulation defines rare diseases as those affecting less than 1 in 2,000 people [6, 7].

Nevertheless, rare diseases are well-recognized to negatively impact patients, their families, and their communities as they are often chronic, progressive, disabling, and life-threatening. Patients with rare diseases may therefore face significant morbidity and

mortality, and a decrease in quality of life. The social and economic burden of rare diseases on the individual and society are also considerable. In a recent systematic review of the direct and indirect costs for ten rare diseases, most of the rare diseases were associated with significant economic burden [8]. For example, the costs of Cystic Fibrosis in Europe can range from €16,307 to €394,518 per patient per year [9–13]. For these reasons, clinical practice guidelines are very useful in supporting clinical decisions, health policy and resource allocation for rare diseases.

Although clinical practice guidelines are important for persons with rare diseases, the guideline development process can be challenging when diseases are rare, which is reflected in the paucity of guidelines for rare diseases. In a single public guideline database, the National Guideline Clearinghouse contains over 2,300 guideline summaries alone. However, the database is comprised mostly of guidelines for common diseases (e.g., heart failure and diabetes) and lacks representation of rare diseases (e.g., hemophilia and cystic fibrosis). As an example, there are currently 47 guidelines on heart failure, while there are only four guidelines on hemophilia A and B. The disparity between the number of guidelines for common and rare diseases can explained by a variety of sources. In 2012, a RARE-Bestpractices Guidelines open forum workshop was held, participants discussed perceived characteristics that set rare disorders apart from common diseases that can either serve as a facilitator or barrier to clinical care, research, and guideline development in rare diseases (Table 1) [14]. Each phase is interrelated – a barrier in the clinical phase may influence the research phase, and a facilitator in the research may influence the guideline development phase, and so forth. As an example of research

influencing guideline development, research in the field of rare diseases tends to be based on observational studies and therefore guideline developers typically rate the quality of evidence as very low to low rather than moderate or high. Additionally, the lack of randomized controlled trials in rare disease research may make it difficult to make strong guideline recommendations.

In response to the apparent barriers to developing guidelines for rare diseases, the four-year (January 2013 – December 2016) RARE-Bestpractices project was developed. RARE-Bestpractices focuses on the best practice and knowledge sharing in the clinical management of rare diseases [15]. One of the main goals of the project is to create transparent and rigorous methodology for the development of guidelines on rare diseases [16]. Once the methodology has been standardized, it can be used by guideline developers to create high quality rare disease guidelines. In October 2013, RARE-Bestpractices held a workshop in Freiburg to form a discussion around potential methodologies and challenges in rare disease guideline development [14]. As a follow-up to the workshop in Freiburg, RARE-Bestpractices seeks to apply the suggested methodologies and identify rare disease guideline development challenges in a pilot project.

1.2 Development of clinical practice guidelines for rare diseases: an example from hemophilia

In May 2012, the National Hemophilia Foundation (NHF) partnered with McMaster University to develop a guideline on care models for hemophilia management in response to the NHF strategic summit, where an increased emphasis was placed on

evidence-based care for persons with hemophilia [17]. The decision to produce an evidence-based guideline on care models for the management of hemophilia presented an opportunity to identify methodologies and challenges for rare disease guideline development.

1.2.1 Guideline development rationale

1.2.1.1 What is hemophilia and why it is considered a complex disease

Hemophilia is an X-linked congenital lifelong bleeding disorder due to mutations in clotting factor genes: factor VIII in hemophilia A; and factor IX in hemophilia B. Hemophilia A and B are rare diseases, with a prevalence of 1 in 10,000 and 1 in 50,000 males, respectively [18, 19]. Care for persons with hemophilia is often complex and requires management beyond the prevention and treatment of bleeding. Consultation with knowledgeable experts, such as hematologists and specialized nurses, can be important to develop hemophilia-specific treatment and management plans, as well as other supportive measures. Health care providers focused on musculoskeletal health, such as physical therapists, may be required to manage bleeding into joints and chronic joint damage [20]. Psychosocial support may be required as persons with hemophilia can experience limitations on their activities, which can result in social stigma and vocational challenges, and decreased quality of life and life satisfaction [21, 22]. These individuals can be supported by social workers and psychologists. Other supplemental support can include infectious disease specialists to manage viral infections acquired from blood products used in the 1980s, gastroenterologists to manage liver disease associated with hepatitis,

dentists for dentition-related complications, and molecular geneticists for carrier detection and prenatal diagnosis. As a result, care for persons with hemophilia is often multidisciplinary and specialized.

1.2.1.2 Care models in the management of hemophilia

The integrated care model

A model of care that is multidisciplinary and specialized can be defined as "integrated care" and is one of the four models currently operating in the US. The integrated care model is represented in the US by the federally funded Hemophilia Treatment Centers (HTCs). Care includes supervision of persons with hemophilia via a coordinated and centralized multidisciplinary team, which can be composed of a hematologist, specialized nurse, physical therapist and a social worker. Another aspect of the integrated care model is home-based treatment, which allows HTCs to help persons with hemophilia self-infuse factor concentrate to treat their bleeds at home. HTCs coordinate care, secure and administer funding, provide technical assistance, organize professional education and training, and engage in data collection and analysis.

The specialist based care in a non-specialized setting model

The specialist based care model centers on a hematologist, who may or may not have specialized training in hemophilia, providing care in a non-specialized center, such as a hospital or medical office.

The non-specialist based care in a non-specialized setting model

Care delivered by a non-specialist in a non-specialist setting can include a family physician delivering care in their practice, or an emergency room physician delivering care in a hospital.

The "no care" model

In the last known care model, the "no care" model, there is a complete absence of care for persons with hemophilia who not have access to care due to profound resource constraints. Figure 1 shows the four observed care models in the US.

Of the four care models that currently exist in the US, there has been strong advocacy for the integrated care model since its development in the late 1940s [23–26] and increased uptake in the 1960s and 1970s. Although the integrated care model is widely accepted today, there are few studies that assess the impact of integrated care in relation to the other existing care models.

1.2.1 Guideline development process and initial stages

The development of an evidence-based, rigorous, and transparent clinical practice guideline involves several steps. The complete guideline development process spans from planning and formulation to reporting and peer review, and eventual updating. Figure 2 shows a simplified depiction of the guideline development process.

This section briefly reviews the initial stages of our guideline development process, which provides a background and context for this thesis work. For a more

detailed description of the guideline development process, refer to the complete guideline and methods papers [27, 28]. In 2013, the initial steps of priority setting; and target audience and topic selection, were conducted. Priority setting involved the identification, balancing, and ranking of priorities by key stakeholders. The NHF organized a committee of stakeholders consisting of current and past chairs of NHF's Medical and Scientific Advisory Council (MASAC), members of the World Health Federation (WHF), and members and consultants of NHF. From a list of priority topics derived from the committee, the topic of care models in the management of hemophilia was chosen to be addressed in the first clinical practice guideline. For this guideline, the foremost target audience was determined to be persons living with or affected by hemophilia. Additionally, the guideline was intended to be a resource for hospitals and health care systems, federal and state programs and policy makers, private and public insurers, and other healthcare professionals developing and choosing strategies to care for persons with hemophilia and other bleeding disorders.

The next step of the guideline development process was question generation. This involved defining key questions the intended recommendations should address that would be relevant for decision-making. At this stage, key stakeholders of MASAC members, US HTC staff, NHF Chapter Presidents, and members of the NHF Nursing, Physical Therapy, and Social Work Groups developed a list of specific questions and outcomes to be addressed. The list of questions and outcomes were reviewed by a guideline panel. The panel consisted of US and non-US healthcare providers with expertise in hemophilia care (physicians, nurses, physical therapists, a genetic counsellor); individuals with experience

in health policy, healthcare financing, and research related to hemophilia; and persons with hemophilia or other rare diseases, and parents of persons with hemophilia. In June 2014, the panel generated the final questions and selected the outcomes rated critical to be addressed in the guideline. It was determined that the first guideline question was to address the impact of different models of care, including the integrated care model typified by US HTCs, on patient-important outcomes. The second guideline question focused on determining the components of the integrated care model necessary to produce improved patient-important outcomes. The two research questions in the population, intervention, comparator, and outcomes (PICO) format are shown in Table 2 and 3.

1.2.2 The guideline development team

To facilitate the development of the NHF-McMaster Guideline on Care Models for Hemophilia Management, a core methods group from the Department of Clinical Epidemiology & Biostatistics, McMaster University, was formed. The group consisted of a guideline methodologist from the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) approach (Dr. Nancy Santesso, NS), a hematologist and laboratory physician with content expertise (Dr. Menaka Pai, MP), an experienced qualitative researcher (Shannon J Lane, SJL), and myself (Cindy HT Yeung, CHTY), a Masters student in the Health Research Methodology program. Our team was led by clinicians and methodology experts, Dr. Alfonso Iorio (AI) and Dr. Holger J Schünemann (HJS). The core methods group met once a week for strategic planning, decision making, and updating one another on our progress. I played a key role in

planning, organizing, and leading these meetings. Overall guidance throughout the guideline development process was provided by AI and HJS.

1.3 Thesis outline

1.3.1 Thesis overview and rationale

The contents of this thesis will focus on the subsequent stages of our guideline development, following the initial phases as discussed in section 1.2.1 of this chapter (Figure 2). The subsequent stages include summarizing evidence and considering additional information; and judging quality (also known as strength or certainty) of the body of evidence. It also includes, steps for determining the effects of the interventions; importance of outcomes and interventions, values, preferences, and utilities; and baseline risk, burden of disease, resource use, effects on equity, and other information. We used the GRADE approach as a rigorous and transparent evidence-based approach to develop this guideline.

Systematic review of care models for hemophilia management

In the standard guideline development process, a literature search is performed and evidence is synthesized, focusing on the effects of the intervention (or diagnostic test); patient values and preferences; and factors on the resource use, equity, acceptability, and feasibility of treatment (or diagnostic test) implementation. In April 2015, we completed this literature search for hemophilia and care models. Continuing with the guideline development process, we would then move to summarizing this evidence and

judging its quality in an evidence table. At the next stage of the guideline development process, an evidence profile and evidence to decision table were created and presented to the guideline panel [29–31]. The evidence profile is a type of evidence table that presents the best available body of evidence related to the health care question for all patient-important outcomes for the panel to derive their recommendations. The evidence to decision table is used along with the evidence profile to facilitate decision making, record judgements, and document the process of going from evidence to the decision. The table includes evidence from the evidence profile of benefits and harms, and factors on resource use, equity, acceptability, and feasibility.

Chapter 2 of this thesis focuses on the systematic review and assessment of the review's evidence in an evidence profile. The chapter also includes a commentary of the many methodological challenges and the solutions we applied. The challenges reflect the difficulties of conducting a systematic review and assessing evidence in the field of rare diseases and healthcare services.

Developing an overview of reviews regarding integrated multidisciplinary care for the management of chronic conditions in adults

The guideline development process normally proceeds with the aforementioned methodology, however the evidence gathering stage is particularly difficult for rare diseases for reasons discussed in section 1.1 of this chapter. Indeed, after conducting the systematic review of hemophilia management and care models, we found few studies and they were of low to very low quality evidence. Rather than stopping at the paucity of

evidence retrieved from the hemophilia and care models literature search, the core methods group decided to use additional methods to gather evidence. Three strategies were chosen to address the lack of evidence from hemophilia and care models: a literature search in chronic conditions and care models; gathering expert-based evidence with systematic observations; and conducting qualitative interviews with key stakeholders. These additional methods gathered evidence that was then summarized and judged for quality, strength and uncertainty, and considered with the hemophilia and care models evidence. Figure 3 summarizes the steps that were completed for our guideline and where these additional methods were used.

Chapter 3 focuses on the first strategy, the literature search in chronic diseases and care models. The process of finding indirect evidence from other chronic conditions led to the objective of conducting an overview of reviews. The overview of reviews examined the effects of integrated multidisciplinary care for the management of chronic conditions in adults. From the process of the overview we gathered evidence for the NHF-McMaster Guideline on Care Models for Hemophilia Management. The evidence from the overview was then assessed using GRADE methodology. As with chapter 2, this chapter also includes a commentary of the many methodological challenges and the solutions we applied. The challenges reflect the difficulties of conducting an overview of reviews and assessing evidence in chronic conditions and healthcare services.

The final stages of evidence gathering for the guideline included other methods of additional evidence retrieval: expert-based evidence by systematic observations and qualitative interviews with key stakeholders (Figure 3). These studies were led by

methods group members, MP and SJL, respectively. For a detailed description of the methodology and results of these studies, please refer to the methodology paper [28] and qualitative study paper [32].

1.3.2 Thesis objective

The objective of this thesis is to identify methodological challenges and facilitators, and the decision-making process to derive solutions to these challenges in the development of a guideline in a rare disease. By discussing the process of developing the NHF-McMaster Guideline on Care Models for Hemophilia Management transparently, this thesis aims to provide a framework for future guidelines in rare diseases.

Chapter 2. Developing a systematic review of care models for hemophilia management

2.1 Introduction

The development of a systematic review on care models in the management of hemophilia presented many challenges. Some of these challenges were expected and inherent in the study of rare diseases and the assessment of health care programs, while others were not anticipated. The outline of this chapter follows the format of the typical steps to conduct a systematic review, from the stage of searching for studies to presenting the results. Methodological challenges within each step of the review development will be discussed, and the accompanying solutions that were applied to overcome these challenges.

2.2 Objectives

The objectives of discussing the development of this systematic review are:

1. To transparently present the decision-making steps in the review process;
2. To identify issues that may arise in conducting a systematic review due to a paucity of high quality evidence using an example from hemophilia; and
3. To identify solutions to methodological challenges, offering guidance for future reviews in rare diseases and health care services.

2.3 Methodology, challenges, and solutions

2.3.1 Searching for studies

In order to begin a systematic review, the sources to search for relevant studies must first be identified. As the McMaster Core Methods Group, we decided that the MEDLINE, EMBASE, and CINAHL bibliographic databases were to be searched. The Cochrane Central Register of Controlled Trials (CENTRAL) was not included as we anticipated that no randomized studies in hemophilia and care models would be found. Grey literature to be searched included previous guidelines in hemophilia such as, the Canadian Comprehensive Care Standards for Hemophilia and Other Inherited Bleeding Disorders by the Canadian Hemophilia Standards Group [33]. For example, references from this guideline were thought to be useful since there were sections outlining guidance on the delivery of comprehensive care, albeit, based mostly on expert opinion and experiences. Snowballing from expert-identified reviews [34–37] and studies [38–40] on hemophilia care models were also seen as a valuable source of search terms and further references.

2.3.1.1 Designing the search strategy

In the first step to design the search strategy we defined search limits. In the case of a rare disease, it was anticipated that not many studies on the topic of care models in hemophilia would be retrieved. Therefore, to find more studies the search strategy was not limited by year, language, or study design. Next, the structure of the search strategy was designed. The following general structure was used: terms to search for the health condition of interest; terms to search for the intervention(s) evaluated; and terms to search for the types of study design to be included [41]. Using this structure, it was determined

that terms for the health condition or population of interest would be related to hemophilia; and the interventions evaluated were the four care models of interest. Restrictions on study design were not used.

The selection of key words from studies or reviews that involved hemophilia patients was performed to retrieve search terms for hemophilia. However, a methodological challenge arose when identifying search terms for care models in hemophilia. Since the integrated care model is widely accepted today, and there was strong advocacy and uptake for it in the 1970s, the remaining three care models have been rarely studied. With few studies in hemophilia based on the other care models to extract intervention terms from, another approach was taken. We decided to use papers from other chronic conditions with established and frequently studied models of care. For example, a Cochrane systematic review on integrated disease management interventions for patients with chronic obstructive pulmonary disease (COPD) was used to define synonyms and related search terms for integrated care and the other care models [42]. Studies of these reviews typically defined the other care models as "usual care" or "no intervention", as their control groups. We found common key words that described the control groups, for instance, care in a "community hospital", "ambulatory care", and care from a "general practitioner only". Another valuable source of terms to describe the other care models was through experts in the hemophilia field, including members from the panel and the Core Methods Group (AI, MP).

Methodological challenge: The concept of non-integrated care, which included the care models: specialist in a non-specialized setting, non-specialist in a non-specialized setting,

and no care, were rarely studied in hemophilia. Therefore, it was difficult to build a search strategy to retrieve these studies.

Solution: We used key words to describe the non-integrated care models from other sources. Other sources included reviews and studies from other chronic conditions with established and frequently studied care models. Additionally, we received input from experts in the hemophilia field to provide further descriptive terms for non-integrated care.

After developing the search strategy with the population and care model terms, we tested the ability of the strategy to detect targeted studies. Although we were successful in retrieving these studies, the Core Methods Group raised a question of whether to include key integrated care terms used in hemophilia, "comprehensive care center" and "hemophilia treatment center". The methodological challenge that arose was whether to increase sensitivity at the expense of precision. The consequence of increasing sensitivity (or comprehensiveness) of a search would be the reduction in precision, thereby retrieving more non-relevant articles [41]. Figure 4 depicts this increase in sensitivity and potential reduction in precision. The solid lines and dotted line between hemophilia and specific care model terms show the relevant records that would be retrieved by including hemophilia "and", delivery of care "or" specific care model terms. As a result, we would retrieve relevant studies that are related to hemophilia and care delivery or our specific care models. The solid line between hemophilia and specific care model terms show the additional records that would be retrieved by including the care center terms, "comprehensive care center" and "hemophilia treatment center". Since a large majority of

studies in hemophilia, regardless of whether they are studying care models, were set in a hemophilia care center (or comprehensive care center), many articles use these key words in their papers. Consequently, it was possible to retrieve an abundance of studies that were not focusing on care models, but had only conducted their study at a HTC. For example, after adding these new terms to the search, we retrieved a study describing the health-related quality of life of persons with hemophilia and the associations between self-reported joint pain, motion limitation, and clinically evaluated joint range of motion. While this study was conducted in persons attending a HTC and compare the results with healthy individuals in the US, the focus was not on the effects of integrated care. Nonetheless, we included this study since it provided single arm quality of life estimates in hemophilia patients attending an integrated care model. Although the reason to include this particular study was clear, for other studies there was more ambiguity and an increased chance of including non-relevant articles. Additionally, the search could potentially become biased towards integrated care as a consequence of increasing its number of terms, while the three remaining models would be underrepresented.

Despite these limitations, we decided to use these terms in our search. By maximizing potentially relevant studies on hemophilia and care models, we were able to retrieve studies that may have not been detected [43–45]. This was because "comprehensive care center" and "hemophilia treatment center" were two of the most commonly used words to describe integrated care in hemophilia. We do however, acknowledge that although we applied this solution, there were limitations to this approach as discussed.

Methodological challenge: Since the concept of integrated care is established and frequently used, should common integrated care terms that may introduce bias be included?

Solution: The care terms, "comprehensive care center" and "hemophilia treatment center" were used. The increase in sensitivity at the expense of precision (potentially increasing the number of non-relevant studies and ambiguity in inclusion criteria, and risking an overrepresentation of integrated care), was judged to be acceptable since it retrieved key articles relevant to the systematic review.

The final search strategy included terms describing hemophilia (factor VIII; factor IX; hemophilia A; hemophilia; hemophilia carrier; etc.) and, delivery of care (patient care planning; patient care management; health services administration; etc.) or specific care models (integrated care; patient care team; comprehensive health care; comprehensive care center; treatment center; home care services; community hospitals; ambulatory care; community health center; etc.). The complete search strategy in each database is presented in Appendix A.

2.3.2 Selection of studies and collecting data

2.3.2.1 Results from the search and records for screening

Our search strategy retrieved 6,788 records after duplicates were removed. These records were available for title and abstract screening. Prior to screening, we wanted to ensure that as much evidence regarding hemophilia and care models would be retrieved as possible. The paucity of studies in the literature was a methodological challenge when

searching for studies on a rare disease (hemophilia) and an intervention that is not commonly studied (care models). First, to collect as much evidence as possible, we included articles of all languages and contacted translators when non-English articles were available. Furthermore, since we included English-written articles on the effects of integrated care when the model uptake was increasing in the 1960s and 1970s, we did not want to exclude studies from other countries such as Russia, Spain, and France, as their data would be equally as useful. Second, since many of these articles were published in the 1970s and 1980s, they were difficult to retrieve online. To ensure they could be appropriately screened for inclusion, physical copies were retrieved from libraries. Third, the search retrieved abstracts from recent conferences that we contacted the authors for full-text if they were available. Lastly, a valuable source to ensure we did not miss any articles was by contacting the panel members, as they were experts who were up-to-date in the hemophilia care literature. This process retrieved three additional studies [46–48].

Methodological challenge: It was anticipated that we would retrieve limited studies on care models in hemophilia.

Solution: Our strategies included, not excluding articles based on language by translating their text; retrieving physical copies of older articles; contacting study authors for available full-text of abstracts; and asking panel members who were experts in the hemophilia field to identify any articles that may have been missed.

Prior to and during the screening process, we found many articles reporting results on the same study. This was mostly due to older studies from the 1960s and 1970s that tended to write correspondences to one another regarding the start of a new therapy or

mode of delivering care. Authors also tended to frequently publish these results to update one another. Identifying multiple reports from the same study could be a methodological challenge for all systematic reviews. We used some of the strategies as outlined by the Cochrane Handbook for useful criteria to compare reports: author names; location and setting; specific details of the intervention (e.g., dose, frequency); numbers of participants and baseline data; and date and duration of the study [41]. For our review, the most useful identifying component was the location and setting, where almost all of the screened studies included the HTC (or any other integrated care center) name and location. For example, a study reporting three years' experience in haemophilia home treatment was similar to a correspondence paper of the same author and year of publication [49, 50]. Since both reports were also from the same location and setting, the University of Pisa, Santa Chiara Hospital in Pisa, Italy, we were more confident that the correspondence paper was supplement to the three-year study.

Methodological challenge: Authors tended to publish the results of their study in multiple reports, particularly studies published in the 1960s and 1970s. It would not be appropriate to present and analyze duplicate data from the same study, however there was difficulty in determining the original and supplementary reports.

Solution: We used strategies as suggested by the Cochrane Handbook, with the study location and setting being the most valuable identifying factor.

2.3.2.2 Developing the screening criteria

Two investigators (SM and BY) independently screened the title, abstract and full text of relevant articles for inclusion. A third investigator (TN or CHTY) resolved all disagreements.

The initial screening inclusion criteria that was used is as follows:

1. Does the study report on individuals with hemophilia or individuals who are carriers of hemophilia? Yes or no
2. Is the main focus of this study to report on specific health care models? The care models include:
    a. Integrated care model (e.g., integrated treatment center, comprehensive care center, team approach, comprehensive care management, comprehensive case management)
    b. Care in a non-specialized center, delivered by a specialist (e.g., community hospital)
    c. Care in a non-specialized center delivered by a non-specialist (e.g., family physician, general practitioner)
    d. No care (e.g., no supportive care)

    Yes or no
3. Does the study provide information on any of the following outcomes?
    a. Mortality
    b. Missed days of school or work
    c. Emergency visits

d. Length of in-patient stay
e. Quality of life
f. Joint damage or joint disease
g. Educational attainment
h. Patient adherence
i. Patient knowledge
j. Caregiver burden
k. Costs
l. Patient values and preferences or other factors (equitable access, acceptability, and feasibility)

Yes or no

All three inclusion criteria had to be met to for the study to qualify for inclusion.

The above screening criteria was developed from the guideline questions in Table 1 and 2. To summarize, the articles had to report on individuals with hemophilia or carriers of hemophilia, while the main focus had to be on a care model of interest, and had to provide information on at least one of our outcomes of interest (patient-important outcomes and those that were necessary for the evidence to decision table in making recommendations).

2.3.2.3 Categorizing studies that focus on care models

The methodological challenge that arose when using the initial screening criteria, was the difficulty in deciding whether the main focus of the study was to report on one of

our specific health care models. As a solution, we used a handful of screened articles to develop a clear care model classification system (Figure 5). First, our system identified whether the focus of the study was classified as either comparative or non-comparative. Second, the comparative studies must have compared outcomes in people who:

1. Received one model of care versus people who received another model of care; or
2. Had "usual care" and subsequently received a model of care (also described as a "pre- and post-intervention" or "before-after" study); or
3. Received integrated care with one set of health care providers versus another set of health care providers (for guideline question 2, see Table 3); or
4. Received one model of care attending at one frequency versus another frequency.

Third, the non-comparative studies must have described a model of care without any comparisons, and report on people who received a model of care that may have an add-on care delivery option (e.g., home care, telehealth) and report an outcome of interest. Table 4 presents examples of retrieved studies that could be organized into each category. There were no identified studies in hemophilia that could be described as the third category of comparative studies: receiving integrated care with one set of heath care providers versus another set of health care providers. This classification system helped to reduce ambiguity in whether to include or exclude studies. As a result, the final inclusion criteria specified that articles were included if they reported on individuals with hemophilia or carriers of hemophilia; the main focus was on a care model of interest (as classified by the hemophilia care model study categories); and provided information on at least one of our outcomes of interest.

Methodological challenge: The inclusion criteria that the main focus of the article was a specific health care model, was not well-defined and therefore created ambiguity in whether to include or exclude articles.

Solution: We developed a classification system to categorize the different main focuses of the articles. If the article did not belong in any of the categories, it was excluded.

2.3.2.4 Data extraction and using results from studies with different objectives

The next step of the systematic review was data extraction. For this step, standardized electronic data extraction forms were developed and pilot tested. Two investigators (SM and BY) independently performed the data extraction, and a third investigator (TN or CHTY) adjudicated all discrepancies.

Methodological challenges arose during the development of the extraction form. Systematic reviews that include only one study design, such as randomized controlled trials, could simply extract data from the treatment and control arms, and calculate a relative and absolute effect. However, it was difficult to decide which data is relevant to review due to the different ways of studying hemophilia and care models. These studies tended to have variability in their focus and thus they had different methodological designs and methods of reporting. For example, one study compared participants with different frequency of visits to a HTC for the outcome, missed days of school or work [45]. Other studies focused on comparing participants before and after the implementation of integrated care and home care [51–53]. By continuing to use our classification system,

we were able to organize the study types and decide which extracted data would belong to each study arm.

For studies that compared outcomes in people who received one care model versus people who received another care model, we used data from the two study arms. For example, a study of 2,950 patients with hemophilia from six states of the US reported outcomes in patients receiving care from HTCs compared to those primarily receiving care from non-HTCs (e.g., private physicians, hematologists, nonhospital-based clinics) [54]. Outcome data from those receiving care from HTCs and those from non-HTCs were both extracted. For studies that compared outcomes in those who had "usual care" and subsequently received a care model, data were used from the before and after intervention arms. As an example, a study of hemophilia patients from 11 federally funded Comprehensive Hemophilia Centers had reported outcomes in 4,742 patients in their fifth year following the initiation of the program, and 2,112 patients in the year preceding the program [52]. For studies that compared outcomes in patients receiving one model of care at different frequencies, data was used from frequent and infrequent model of care users. A study of 6,420 hemophilia patients from 130 HTCs in the US provided an example of this type of study [45]. Outcome data from patients who were frequent users (one or more HTC visits per year) and data from infrequent users (less than one visit per year, excluding the first visit) were both used.

Two types of studies described a care model that provided comparisons between the presence and absence of an add-on care delivery option, rather than comparing care models. Add-on care delivery options could be described as not necessary components of

a care model, but acting in addition to an existing care model to improve the delivery of care. Common add-on care delivery options include home care (such as factor concentrate infusions at the home of hemophilia patients) and telehealth services offered by HTCs. For example, a study of 45 pediatric patients from a single medical center studied patients before and after the introduction of a 12-month home infusion program [55]. Outcome data was used from only the "after" intervention (home infusion program) arm to describe the integrated care model. These data were used since add-on care delivery options were seen as standard of care for most integrated care models. The second type of study under the non-comparative descriptive care model category, were studies that provided single-arm study data only. For example, a study of 166 patients with congenital bleeding disorders (hemophilia, von Willebrand disease) at a single medical center were followed for one year [56]. The outcome data from all patients in the integrated care model during the one year of follow-up was used.

Methodological challenge: It was difficult to extract and decide how the data were to be used from studies with different objectives to study care models in hemophilia.

Solution: We used our previously developed classification system and extracted data from the included studies according to their category.

Despite developing a strategy to extract the data from the studies, it was overall difficult to retrieve data on our subgroups and outcomes of interest. After adjudicating a few studies with discrepancies between screeners, it was noticed that outcomes of interest that were reported in the articles were usually missed. These studies tended to report outcomes that were not of their interest, descriptively in the results section or in the

discussion. For example, a study by Isarangkura *et al*. in 1987 [57] studied 10 pediatric patients with hemophilia in Thailand. Their main outcomes included number of organ sites of bleeding, total episodes of bleeding, and usage of blood products per year. However, if the article was not read thoroughly, a sentence in the results section regarding the death of one patient would have been missed. To ensure we did not exclude studies reporting our outcomes of interest, we trained our screeners to be more inclusive for title and abstract, and to look for the reporting of outcomes narratively in the text when looking at full-text.

Methodological challenge: While some studies reported their main outcomes of interest clearly, our outcomes of interest were commonly presented narratively in the text and not in the result section. Since there is a paucity of studies on hemophilia care models, it was essential that we minimized the risk of missing relevant studies.

Solution: We trained our screeners to be more inclusive at the title and abstract stage, and be more thorough at the full-text screen to look for narratively reported outcomes.

After data extraction, we presented the results as risk ratios (RR) or mean differences (MD) with corresponding 95% confidence intervals (95% CI), or presented narratively. We pooled the results when the estimates were clinically homogenous and the outcome measure were appropriate to be combined (e.g., standard deviations were provided and sufficiently similar quality of life measures were used). For studies that described a single model of care and had available data, either the risk of an event (or proportion) was calculated, or the mean (or median) and standard deviation of the outcome point estimate were calculated.

2.3.3 Assessing quality of evidence in included studies

For the comparative studies, data were synthesized by pooling estimates where appropriate. For the non-comparative studies, we did not pool the means (or medians) or proportions. The next step of the systematic review process was to assess the quality of evidence.

We used the GRADE approach, two investigators (MP and AI) evaluated the quality of the evidence for each outcome, and a third investigator resolved any discrepancies (NS or CHTY) [58]. The quality of the evidence was assessed as high, moderate, low, or very low. The effect estimates or narrative summaries, and quality of evidence were summarized in a GRADE evidence profile [59].

Since all of the included studies were observational in methodology, evidence started at low quality and could be downgraded based on the GRADE domains: risk of bias, inconsistency and imprecision of the effect, indirectness, and publication bias [60–64]. The studies could also be upgraded based on other criteria as discussed in the sections below.

Risk of bias

According to GRADE, systematic reviews of tools to assess the methodological quality of non-randomized studies (observational studies) have identified more than 200 checklists and instruments [65–68]. To assess the risk of bias in the studies included in the systematic review, it was necessary to use a tool that could appraise different

observational study designs, including cross-sectional, before-after, and cohort studies. The tool should also account for potential important biases in the study of hemophilia and care models. These could include, bias of selection of participants into the study, departures from intended interventions, failure to adequately control confounding (e.g. failure to match for prognostic factors and/or lack of adjustment in statistical analysis), differential missing data, and incomplete follow-up [60, 69].

We decided to use the ACROBAT-NRSI (A Cochrane Risk of Bias Assessment Tool: for Non-Randomized Studies of Interventions) [69], which has recently been updated as The Risk Of Bias In Non-randomized Studies – of Interventions (ROBINS-I) assessment tool [70]. The ACROBAT-NRSI tool appeared to meet the requirements necessary for assessing our included studies in hemophilia and care models. The tool provided detailed guidance for assessing risk of bias, including: bias due to confounding, bias in selection of participants into the study, bias in measurement of interventions, bias due to departures from intended interventions, bias due to missing data, bias in measurement of outcomes, and bias in selection of the reported result.

Methodological challenge: With over 200 checklists and instruments to assess risk of bias in observational studies, an appropriate assessment tool had to be chosen for use.

Solution: We chose to use the ACROBAT-NRSI with the tool being appropriate for assessing different observational study designs (including cross-sectional studies without controls) and addressing potential risk of bias factors important in the study of care models in hemophilia.

Overall, the common biases that the studies had were due to confounding, bias in the selection of participants into the study, and bias due to missing data. Bias due to confounding was common in cross-sectional and before-after studies. For example, Lazerson [53] conducted a before-after study with 20 hemophilia patients who were compared prior to and after the development of a comprehensive care center. The investigators measured missed days from school or work by the number of days lost from school or work, however, no adjustments in the analyses were performed. Potentially important factors that the authors could have adjusted for were that some patients had hepatitis and some were not on prophylaxis treatment. Another before-after study by Smith *et al.* [51] provided an example of bias in the selection of participants in the study. The study included 23 participants before and 43 participants after the development of a Comprehensive Hemophilia Center. There was moderate risk of bias from potential confounding due to the selection of participants into the study. Bias could have been introduced if selection into the study was related to the intervention (integrated care) and the outcome (number of visits to the emergency room and walk-in clinic). It is possible that the 40 patients who chose not to join the program could have had important prognostic factors (insurance status, less severe form of the disease, socioeconomic status, lack of access to the care center, etc.) that were systematically different from the study participants, and thus bias the study estimate for number of emergency room and walk-in clinic visits. The authors did not provide reasons for their non-participation. The study by Lazerson *et al.* [53] provided an example of bias due to missing data. The results of the study showed that 30% of data from patients before the initiation of integrated care was

missing. The result of this selective missing data were that the mean days of absence per school year prior to integrated care could have been underestimated, thereby biasing the effect towards the null.

Table 6 shows a summary of the risk of bias assessment by the ACROBAT-NRSI tool for the eight comparative non-randomized studies.

Inconsistency

According to the GRADE approach, there were several criteria for evaluating the consistency across studies: similarity of point estimates, the extent of overlap of CIs, and statistical criteria [62]. For randomized studies, statistical criteria includes tests of heterogeneity, such as the I^2, which quantifies the proportion of variation in point estimates due to among-study differences [62]. *A priori* hypotheses on expected sources of heterogeneity should be explored in subgroup analyses. For our systematic review, we would expect heterogeneity in persons with hemophilia due to severity, comorbidities (such as HIV/AIDs and hepatitis C), and inhibitor status. We would also expect heterogeneity in the intervention, such as how the care model was delivered. There could also be heterogeneity within care models. For example, there could be different compositions of integrated care due to the number and type of health care providers in the multidisciplinary team, whether add-on care delivery options are present, whether there is organized professional education and training, and other components.

Since we could only assess inconsistency in outcomes with pooled results, only three were assessed: missed days of school or work, emergency room visits, and joint

damage or disease (and other measures of functional status). As these were non-randomized studies, we did not use the I^2 statistic. Instead, we based our assessments from the similarity of point estimates and the extent of overlap of CIs. We rated the evidence from these outcomes as not serious for inconsistency. Additionally, even if there were indications of inconsistency, we would not be able to study the previously outlined potential sources of heterogeneity in the population and intervention due to the lack of subgroup information from these studies.

Indirectness

Evidence could be indirect in one of four ways: differences in population (applicability), differences in interventions (applicability), differences in outcome measures (surrogate outcomes), and indirect comparisons. We did not have any reasons to downgrade the evidence for differences in population since all of the studies included persons with hemophilia. However, there were differences in the study interventions, or applicability of the study interventions to our intervention of interest. The first difference was due to some studies being conducted in the 1960s or 1970s. As a result, three studies had integrated care models delivered in the 1960s and 1970s, and we considered this evidence to be indirect since the treatment modality for hemophilia patients have changed since this time [51–53]. The second difference was due to our definition of the intervention for some studies. For two studies, we used the definition of frequency to describe integrated and non-integrated care [45, 71]. We acknowledged that, using frequent users (attending a HTC with one or more visits per year) and infrequent users

(attending less than one visit per year, excluding the first visit), were indirect measures of integrated and non-integrated care. Another type of indirectness that was assessed was differences in outcome measures, or the use of surrogate outcomes. Since we found few studies assessing emergency room visits, we included a prospective cohort study measuring the number of people with at least one hospitalization over four years. Although the hospitalizations had to be due to haemorrhagic bleeding complications and patients were likely admitted to the emergency setting, the uncertainty that hospitalizations were equivalent to emergency room visits led us to consider some indirectness. No studies were assessed with the last type of indirectness. Indirect comparisons from the lack of head-to-head comparisons (i.e., integrated care versus non-integrated care) were not relevant for the systematic review.

Imprecision

GRADE suggested that the examination of 95% CIs provides the optimal primary approach to decisions regarding imprecision. We considered the sample size, the number of events, and the width of the 95% CIs. Overall, we found imprecision on outcomes with studies that did not calculate or present information on standard deviations [52], and studies with few events [43].

Publication bias and factors for upgrading

Non-randomized studies (observational studies) began as low quality of evidence and could be rated down for the five reasons as previously discussed. However, there

were also three reasons to potentially rate up the quality [58]. GRADE suggested to consider upgrading the quality of evidence when methodologically rigorous observational studies show that a large magnitude of effect exists, a dose-response gradient is present, and all plausible confounders or other biases increase our confidence in the estimated effect [72]. In our assessment of the evidence, we did not find any reason for rating up the quality of evidence.

2.4 Results

We searched MEDLINE (1946 to April 22, 2015), EMBASE (1974 to April 22, 2015), and Cumulative Index to Nursing & Allied Health Literature (CINAHL) (1981 to April 22, 2015) databases. The results of the literature search identified 6,789 non-duplicate records. After title and abstract screening 197 articles were assessed for eligibility with full-text review, and subsequently, 172 articles were excluded. Seven additional articles eligible for inclusion were retrieved during the process. As a result, 27 unique non-randomised studies (which were reported in 32 published articles) were included. A PRISMA diagram of the selection flow is provided in Figure 6. Eight studies were comparative and 19 studies were single arm non-comparative studies.

The summary of the included comparative study characteristics are presented in Table 5 and the characteristics of the non-comparative studies are found in Appendix B. The comparative study characteristics table describes the patients, intervention, controls, and outcomes of each comparative study. We retrieved evidence mainly on the integrated care model. Therefore, all the comparisons were between integrated care and non-

integrated care. The non-integrated care arm described any care model that was not categorized as integrated care. These could be the remaining three care models grouped together, or compared separately with integrated care. Table 5 shows which data were used as the intervention and control groups. We did not find any studies that compared the remaining three care models with one another.

Following the assessment of the quality of evidence from the comparative studies, we had to decide on what evidence to present for the systematic review and guideline to base the recommendations on. We decided to present both the comparative and non-comparative evidence to the panel in order to be transparent about the evidence for hemophilia care models. However, we focused on the comparative studies that were assessed as the best available evidence for decision-making. The results from the comparative studies were presented by outcome in an evidence profile (Table 7) and are described in the following sections. The non-comparative evidence was briefly described as well (Table 8).

Mortality

One comparative non-randomized study by Soucie *et al.* [54] with 2,950 participants, reported results on adjusted mortality over a three-year period. The risk of death was reduced in PWH receiving care at an HTC compared to not receiving care at an HTC (RR 0.60; 95% CI: 0.50 to 0.80), with an absolute overall death rate of 40.4 deaths/1,000 person-years, reflective of the burden of HIV mortality [54]. Six unique non-comparative studies (which were reported in seven published articles) reported

mortality as an event rate [55, 57, 73–77]. The number of deaths in the study population over 1 to 8 years, ranged from 6 to 100 deaths per 1,000 persons (Table 8). Overall, the quality of evidence for a reduction in mortality with integrated care was low based on data from the non-randomised comparative study (Table 7).

Missed days of school or work

Three studies measured days lost from work or school. Two before-after studies measured the mean number of days missed prior to and after implementation of an HTC (Table 7). Lazerson [53] with 20 patients before and after the implementation of integrated care, found a reduction in days [MD -50.20 (95% CI: -61.68 to -38.72) days per year]. Smith and Levine [52] with 2,112 participants before and 4,742 after the adoption of integrated care, found a reduction of 10.2 days per year. The third study [45] compared participants with >11 days lost from work who had frequent HTC use to infrequent or first time use. We re-analyzed the data from 6,420 participants and found a risk ratio of 1.01 (95% CI: 0.75 to 1.36) (Table 7). Seven unique non-comparative studies (which were reported in twelve published articles) reported a range of results with some reporting low rates in absenteeism but others reporting greater than 15 missed days per year [47, 48, 53, 75, 77–84] (Table 8). Overall, there was very low quality evidence for a reduction in missed days of work or school per year of approximately 10 days. This was primarily due to lack of adjustment for confounding factors in the comparative studies from Smith and Levine [52] and our re-analysis of Monahan *et al.* [45], and to the potential for little to no difference in days missed (Table 7).

Emergency room visits

Two comparative non-randomised studies measured the number of emergency room visits or hospitalizations. Soucie *et al.* [86] with 2,546 participants, described the number of people with at least one hospitalization over four years. They reported that persons with hemophilia who had received care at an HTC any time during the study period were hospitalized less than those who did not receive care at an HTC (RR 0.60; 95% CI: 0.50 to 0.70) (Table 7). In agreement with Soucie *et al.* [86], a small study by Smith, Keyes and Forman [51] of 43 participants before and 23 participants after the implementation of integrated care, reported a mean difference of 23.3 emergency room and walk-in clinic visits favouring integrated care (Table 7). Three non-comparative studies reported the mean number of emergency room visits [49, 50, 87] and ranged from 0.9 to 500 per 1,000 persons per year (Table 8). Overall, the quality of evidence from the non-randomized comparative studies was low (Table 7).

Length of in-patient stay

One comparative non-randomized study by Smith and Levine [52] of 2,112 participants before and 4,742 participation after the adoption of integrated care, reported number of days spent in hospital. The mean difference in length of stay after implementation of integration care was -7.6 days spent in the hospital per year (95% CI not reported) (Table 7). Thirteen non-comparative studies reported the mean number of days spent in the hospital per patient per year [55, 74–76, 78, 82, 83, 87–92]. The mean

number of visits ranged from 0.4 to 14.5 per person per year (Table 8). Overall, the quality of evidence was very low from the non-randomized comparative study which had a high risk of bias due to non-adjustment for potential confounding and potential bias from patients lost to follow-up and non-participation. These factors were considered with indirectness (study was from the 1980's and current treatment modalities have changed) and the few hospitalizations that occurred (Table 7).

Joint damage or disease (and other measures of functional status)

Two comparative non-randomized studies reported on the progression of joint damage or decreased activity (another measure of functional status) per year. Soucie *et al.* [71] with 4,343 participants, found that for persons with hemophilia with severe disease, frequent HTC users (one or more visits per year) had less range-of-motion (ROM) limitation than infrequent users (less than one visit per year). In contrast, for persons with hemophilia with moderate and mild disease, frequent HTC use was associated with higher ROM limitation, even when the association was tested in a model adjusted for age and BMI. The authors appropriately suggest that for mild and moderate patients, frequent bleeds may drive both frequency of HTC visits and limitation in ROM, creating a spurious (confounded) association (Table 7). From Monahan *et al.* [45], we re-analyzed the data from 6,420 participants to compare participants with decreased activity (actions related to work, school, recreation, and self-care) who had frequent HTC use to infrequent or first time use and calculated RR 1.20 (95% CI: 0.98 to 1.46). Two unique non-comparative studies (which were reported in three published articles) reported [57, 79,

80] a range from 234 to 333 joints damaged or diseased per 1,000 persons (Table 8). Four non-comparative studies (which were reported in six published articles) reported [55, 78, 81, 84, 85, 93] that the proportion of patients with joint damage or disease ranged from 44 to 429 per 1,000 persons (Table 8). Overall, there was very low quality evidence for a reduction in joint damage or disease from the two comparative studies due to lack of adjustment for confounding factors in combination with indirectness (using decreased activity per year to define joint damage or disease) (Table 7).

Patient knowledge

The comparative non-randomized study by Arnold *et al.* [43] of 104 participants found that HTC attendance within the past 12 months was significantly associated ($p<0.05$) with increased knowledge seeking (e.g., recognizing and treating a bleed, knowledge of the genetics of hemophilia, physical activity selections) in an unadjusted analysis (Table 7). The overall quality of evidence was very low due to potential bias from unadjusted confounding factors, indirectness from defining integrated care as attendance at an HTC in the last 12 months, and few participants (Table 7).

Other outcomes: quality of life, educational attainment, patient adherence

We did not find comparative studies reporting on quality of life, educational attainment, and patient adherence. However, two non-comparative studies reported that 83 to 494 per 1,000 persons adhered to their treatment regimens [56, 81] (Table 8).

2.5 Conclusion

We found low to very low quality evidence that integrated care reduces mortality, emergency room and walk-in clinic visits, hospitalizations (and length of stay), missed days of school and work, and increases knowledge seeking. The evidence for the effects of integrated care on functional status, measured by joint damage or joint disease was less clear, and the analysis is likely confounded by disease severity. This means that the true effects of integrated care may be substantially different from what we found (i.e., over-estimated or under-estimated). We also did not find evidence to compare the effects of integrated care to other models for quality of life, educational attainment, and patient adherence.

Through the systematic review process, we identified several methodological challenges. Some of these challenges were inherent in the study of rare diseases and health care programs, while others were due to the paucity of evidence in hemophilia care models. These challenges allowed us to develop solutions that can be used as an example for the evidence gathering, synthesis, and appraisal stage of the guideline development process for other rare diseases.

Despite our comprehensive search in electronic databases supplemented with snowballing and broad expert consultation, and other methods to capture as many relevant studies as possible (by expanding the search strategy to include care center terms, not restricting to date and language, etc.) we found few comparative studies yielding a limited body of evidence. As a solution to the lack of high quality studies in hemophilia, we decided to seek higher quality evidence (i.e., randomized controlled trials) from

chronic conditions with established care models. The next chapter describes the process of the evidence gathering, synthesis, and appraisal in other chronic conditions. However, similar to our systematic review in care models for hemophilia management, our process in other chronic conditions was not without methodological challenges and solutions. These will also be discussed in the next chapter.

Chapter 3. Developing an overview of reviews regarding integrated multidisciplinary care for the management of chronic conditions in adults

3.1 Introduction

Following the systematic review on care models in the management of hemophilia, we concluded that there was a paucity of high quality of evidence. Rather than restricting our evidence to low to very low quality from the hemophilia literature, we decided to also use higher quality indirect evidence from other chronic conditions with established care models. The plan *a priori* was to use evidence from other chronic conditions together with the evidence from the systematic review in hemophilia to guide the panel in formulating recommendations.

Although the process of gathering evidence from other chronic diseases and care models was initiated by the development of the guideline, we decided to broaden the scope to report and assess the quality of evidence from all the relevant reviews. As a result, we conducted an overview of reviews on integrated multidisciplinary care for the management of chronic conditions in adults.

Similar to the systematic review on hemophilia care models, the development of the overview of reviews also had methodological challenges. However the challenges differed from the hemophilia review because they were not due to the difficulty of studying hemophilia as a rare disease, its health care services, or the paucity of high quality of evidence. Instead, the methodological challenges in this chapter outline the difficulty of conducting an overview of reviews – a relatively novel approach to studying evidence at the review-level – and the unique process of using this evidence for studying

hemophilia. The decision-making process and the solutions we applied will also be discussed in this chapter.

3.2 Objectives

The objectives of discussing the development of this overview of reviews were:

1. To transparently present the decision-making steps in the overview process;
2. To identify issues that could arise in conducting an overview of reviews and using this evidence for rare diseases, including an example from hemophilia; and
3. To identify solutions to methodological challenges, offering guidance for future overviews of reviews and the use of this evidence for other rare diseases.

3.3 Methodology, challenges, and solutions

3.3.1 Scope of the overview of reviews

Overviews of reviews are intended primarily to summarize multiple intervention reviews. They have a similar structure to intervention reviews, but focus on reviews rather than primary studies. In the section of the Cochrane Handbook, titled "Special topics", the chapter describing the methodology for overviews of reviews outlines several reasons for conducting an overview. The examples include, to summarize evidence from more than one systematic review of different interventions for the same condition or problem; to summarize evidence from more than one systematic review of the same intervention for the same condition or problem where different outcomes are addressed in different systematic reviews; and to summarize evidence about adverse effects of an

intervention from more than one systematic review of use of the intervention for one or more conditions [41]. The most relevant reason that applied to our overview process was to summarize evidence from more than one systematic review of the same intervention for different conditions, problems, or populations.

3.3.1.1 Objectives of the overview of reviews

The first methodological challenge was deciding the objectives of the overview. The initial objective was to focus on how we used evidence from other chronic conditions for the NHF-McMaster Guideline on Care Models for Hemophilia Management. After much deliberation, we decided to expand the scope of the overview after recognizing that our overview involve a two-part process. First, we would use a literature search to identify and retrieve all relevant reviews from chronic diseases with established care models. We would then assess the evidence from all of the included reviews. Second, we would assess the indirectness of the evidence from these reviews to hemophilia for the guideline. Since we had to gather all relevant reviews and assess their quality of evidence in the first part of the overview process, we decided it was appropriate to present these results regardless of whether they were direct to hemophilia. Methodologists and other users studying integrated care models in chronic conditions could be interested in this evidence. Together with an example of how others could use this evidence for rare diseases or other sufficiently direct populations, we formulated two objectives for the overview:

1. To present the results of our appraisal of the evidence on integrated care for chronic conditions
2. To present the use of this evidence in the development of a guideline on a rare disease such as hemophilia

Methodological challenge: The main focus of our overview process was to report on the methods and results of the evidence that contributed to the NHF-McMaster Guideline on Care Models for Hemophilia Management. However, since the overview process covers results and quality of evidence assessments beyond the scope of the guideline, there was potential to broaden the scope of the overview.

Solution: We decided to broaden the scope of the overview of reviews to report on all the relevant reviews in chronic conditions and care models, in addition to the evidence we used for the guideline in hemophilia.

3.3.1.2 Overall search strategy plan

Before designing a search strategy to retrieve evidence on chronic conditions and care models, it was necessary to clarify the general approach we were to take. We first ran the systematic review on hemophilia and care models search strategy with chronic conditions (and synonyms such as chronic disease and illness) replacing the hemophilia terms. The specific chronic conditions were chosen by the guideline panel to have established care models. These included chronic obstructive pulmonary disease (COPD), cystic fibrosis, diabetes mellitus, HIV/AIDS, and heart failure. We also added search terms for randomized controlled trials, quasi-randomized trials, and clinical trials with the

objective to retrieve higher quality of evidence as compared to the observational studies from the hemophilia systematic review. After running the search, we found close to 40,000 articles in a single database (MEDLINE). It became evident that it would not be possible to use this search strategy, as it would not be feasible to screen through the large number of articles. However, the search results were useful since they revealed that there were an abundance of potentially relevant high quality studies in chronic conditions and care models. By the process of snowballing the referenced studies of these studies, we found that many were included in systematic reviews. Through these systematic reviews we found a meta-review on integrated care programmes for adults with chronic conditions by Martínez-González *et al.* [91].

The methodological challenge arose when deciding whether the use the search strategy and included reviews by the meta-review. The first consideration was that if we used the meta-review, we could use their included reviews and simply update the search strategy from March 2012 to present. However, we considered that if we used their included reviews that were decided by their screening criteria and we would always be restricted by their screening decisions. This point is outlined in more detail from reviewing the implications on the population, interventions, and outcomes of interest.

Types of populations

The meta-review by Martínez-González *et al.* [91] included adults with non-communicable chronic conditions and excluded reviews on addiction and mental disorders. This implied that we could not examine the pediatric population, conditions

that were communicable (e.g., HIV/AIDS and hepatitis diseases), and addiction and mental disorders (e.g., depression). Although the included population would be restricted, we also considered that the pediatric population typically receive different care strategies from adults (e.g., disorders starting at birth such as cerebral palsy versus adults with heart failure), and that those with addiction and mental disorders have a delivery of care sufficiently different from other chronic conditions (e.g., focus on the psychological and social aspects of the condition). There is however, an important limitation for the guideline in hemophilia by restricting our population according to the meta-review. We acknowledged that persons with hemophilia could be of any age, with a majority in the pediatric population, and that the psychological and social factors could play a large role in their lives due to stigma and its effect on a person's daily life.

It was also important to consider that the search strategy by the meta-review did not use specific chronic condition terms. We found from our initial search strategy that only including chronic condition terms could miss studies that did not use specific chronic conditions to describe their participants (e.g., terms for COPD and heart failure). We contemplated the importance of increasing the sensitivity of this search. If we were concerned with including as many results as possible, we would not to use the meta-review and instead use our original search strategy with additional limits (e.g., restrict by date and choose only two or three chronic conditions to search). However, we noticed that individual systematic reviews typically included specific chronic condition terms. As long as the systematic reviews were described by chronic condition terms. If chronic condition

terms were used to describe the systematic reviews, it would be possible to find these primary studies from the reviews.

Types of interventions

The meta-review was interested in the effects of integrated care programmes which were defined as the integration of healthcare from the provision of multidisciplinary interventions at different stages of the care process in two or more different institutional areas [94]. Reviews that reported on transition of services and end-of-life care were also included. Since the publication of this meta-review, other definitions and conceptualizations of integrated care had developed. The idea of integrated care is complex and increasingly being recognized. Strandberg-Larsen and Krasnik [94] identified the different types of methods used to measure integrated healthcare delivery and emphasized structural, cultural, and process aspects. They found a total of 24 methods available for measuring integrated care, but noted that there still is no consensus on how to measure the concept. Valentijn *et al.* [95] described integrated care in terms of primary care playing a central role within health systems. They defined integrated care with three dimensions, the macro (system) level, the meso (organisational) level, and the micro (clinical) level. The macro-level focuses on the combination of structure, processes, and techniques to fit the needs of populations across the continuum of care. It focuses on horizontal (peer-based and cross-sectorial collaboration, such as connecting primary care providers together) and vertical integration (integration of care across sectors or specializations, such as primary care connected to secondary and tertiary

care) on a system level to improve the overall health of the population. The meso-level is based on inter-organizational and professional integration. Inter-organizational integration relies on the common governance mechanisms to deliver comprehensive services to a defined population. For example, a partnership of an acute care hospital, a children's rehabilitation hospital, and a home/community health organization focused on children with complex conditions [96]. Professional integration relies on the partnerships between professionals both within and between organizations, similar to the horizontal and vertical integration discussed previously, but on a smaller scale. An example of professional integration is a primary health care provider referral to a tertiary care provider. The micro-level is developed from the shared responsibility between the professional and patient to find a common ground on clinical management. The physician and patient interact to coordinate individual patient care, emphasizing personal needs and values. Integrated systems at all three levels are thought to use the integrative guiding principles of primary care, a person-focused and population-based care.

Using the three dimensions of integrated care, we categorized the meta-review as focusing mainly on the inter-organizational and professional integration aspects of the meso-level. Therefore, it would not be possible to study integration at the macro- or micro-level as the reviews have been screened to include meso-level reviews. For the purpose of our overview, the restriction to meso-level studies would be compatible with our objectives since integrated care for hemophilia patients is usually multidisciplinary (i.e., from inter-organizational and professional integration). The other levels of integration could also influence the care of persons with hemophilia, but would be beyond

the scope of using this evidence for the guideline. Additionally, an overview of reviews based on the meso-level could still hold value beyond its use for the guideline in hemophilia as integrated care for chronic conditions typically involve integration at the meso-level.

Types of outcome measures

The outcomes of interest for the meta-review were patient-centred outcomes, process quality, and use of healthcare resources and costs. Patient-centered outcomes were defined as outcomes from medical care that patients care about, such as survival, function, symptoms, and health-related quality of life. The results of the meta-review were later grouped into six categories: clinical (e.g., morbidity, symptoms, disease control, and mortality), patient-reported (e.g., patient satisfaction, quality of life, health literacy, and patient preferences), functional (e.g., functional status, exercise capacity, and level of disability), process (e.g., adherence to treatment guidelines, treatment compliance, physician behaviour, and contact with services), use of healthcare resources (e.g., hospital admissions and re-admissions, visits to general practitioners or emergency departments), and costs (e.g., direct and indirect costs to patients, payers, or society). In regards to the guideline in hemophilia, the patient-important outcomes of interest could be categorized under the included outcomes of the meta-review.

After considering the advantages and limitations of using the search strategy of the meta-review and the included reviews, we assessed that the meta-review was performed with a

robust approach and reported on populations, interventions, and outcomes of our interest. We chose to update their literature search, assess the newly retrieved reviews, and the reviews previously retrieved from the meta-review.

Methodological challenge: We determined that a search strategy consisting of chronic conditions, specific chronic diseases, care model, and specific study design terms would lead to an unmanageable number of articles to screen. We found a meta-review by Martínez-González *et al.* [91] that could be applied for the overview, however, there were many advantages and limitations to this option.

Solution: After weighing the advantages and limitations of using the meta-review, we decided use the meta-review as it was performed with a robust approach and reported on populations, interventions, and outcomes of our interest.

3.3.1.3 Definition and scope of integrated care

The meta-review by Martínez-González *et al.* [91] used the term "integrated care" to describe their intervention of focus. We acknowledged that the concept of integrated care could be broad. The integration of health care could occur across systems, organizations, and on a clinical-level without explicitly specifying which health care providers were involved. For example, integrated care for persons with hemophilia could be present on the meso-level through inter-organizational integration. In hemophilia, organizations at the federal level work with manufacturers to reduce factor prices for selected organizations (these would be the HTCs that fall under the 340B program [97]). However, the concept of integrated care in the hemophilia field would be care delivered

by a comprehensive care center or HTC, with a main focus on the multidisciplinary nature of the health care team. To clarify our focus of integrated care, we defined the concept as "integrated multidisciplinary care".

Similar to the systematic review on care models in the management of hemophilia, it was difficult to study the non-integrated care models. Firstly, non-integrated care could be defined differently across chronic conditions. For example, a randomized controlled trial retrieved by one of the reviews studied the effect of a two-year interdisciplinary community-based COPD management programme (integrated multidisciplinary care) on 199 patients with COPD [98]. Usual care (non-integrated multidisciplinary care) involved care from their chest physician providing pharmacotherapy according to accepted guidelines and short smoking cessation advice. If nutritionally depleted, the respiratory physician advised the patient to eat more. This comparator group would be best classified as "specialized care from a non-specialized care setting". However, since the usual care group could be defined differently across conditions and were always compared to the integrated multidisciplinary care group, we referred to this arm as non-integrated multidisciplinary care. Therefore, the overview aimed to study the effects of integrated multidisciplinary care on the management of chronic conditions in adults.

3.3.2 Searching for reviews

3.3.2.1 Clarifying the unit of analysis

One methodological challenge we encountered was defining our unit of analysis throughout the overview development process. By definition, an overview is based on reviews as the unit of interest. To clarify the process, we considered the unit of analysis from the literature search to quality of evidence assessment stage. The unit of analysis for the literature search and selection stage were reviews. This entailed using the meta-review search strategy, which would use search terms to identify systematic reviews. These reviews would then be screened by the inclusion criteria for relevant systematic reviews. Risk of bias assessments of the review would also take place on the review level. Until this stage, reviews would be the studied unit. However, since the body of evidence reporting on outcomes from each review would be comprised of primary studies, the focus would shift to primary studies as the unit of analysis when performing quality of evidence assessments.

Methodological challenge: Although by definition an overview of reviews focuses on the review level, it was necessary to clarify the unit of analysis throughout the overview process, from the literature search to the evidence quality assessments.

Solution: We clarified that the literature search, selection stage by inclusion criteria, and risk of bias assessments would focus on reviews as the unit of analysis. The stage of quality of assessments would be performed at the outcome level with primary studies as the unit of analysis.

3.3.2.2 Updating the meta-review

After clarifying the unit of analysis to be reviews for the search process, the next stage was to update the meta-review. The meta-review searched in MEDLINE (1946 – March 2012), EMBASE (1980 – March 2012), CINAHL (1981 – March 2012), and the Cochrane Library of Systematic Reviews (issue 1, 2012). In order to update these results, we searched from 1 January 2012 to 7 January 2016 in all the aforementioned databases. Additional articles were identified through reference lists, key journals, and reviews.

3.3.2.3 Structure of the search strategy

The search strategy by the meta-review consisted of MeSH headings, keywords, and text words related to integrated care and chronic conditions. The searches were not restricted by age, language, or country. For the full search strategy, see the meta-review by Martínez-González *et al.* [91].

3.3.3 Selection of reviews and collecting data

3.3.3.1 Developing the screening criteria

Two investigators (DZ and AW) independently screened the title, abstract, and full text of relevant articles for inclusion. A third investigator (CHTY) adjudicated all discrepancies.

At the stage of review selection, we were challenged with whether to use the inclusion criteria by the meta-review or develop additional criteria. After reviewing the 27 included reviews by Martínez-González *et al.* [91], we concluded that the applied

criteria was too broad for the objectives of the overview. From the meta-review we used the inclusion criteria that reviews had to be systematic and report on adults with any non-communicable chronic conditions excluding addiction and mental disorders. We developed the following four additional criteria.

First, the review had to report on at least one specific chronic condition. Some of the reviews included by the meta-review did not focus on specific chronic condition populations. For example, a review by Higginson *et al.* [99] reported on end of life care, without specifying the specific chronic conditions of the included population. Furthermore, a review by Boult *et al.* [100] reported on chronic disease models of comprehensive care for chronically ill older persons, but did not report specific conditions. Since we intended to present the evidence by outcome of interest and chronic condition, these reviews would not be included as the chronic conditions were not clear.

Second, the review had to include at least one randomized controlled trial reporting on a multidisciplinary team approach with two or more health care providers. To retrieve the highest quality of studies, we specified at least one study had to be a randomized controlled trial. The composition of the intervention health care providers was used as an additional inclusion criterion to ensure at least one study had a team of health care providers delivering care.

Third, the review had to report on at least one of our outcomes of interest in a quantitative (meta-analysis) or narrative summary. Some of the reviews included by the meta-review did not synthesize the results of our outcomes of interest from the included studies. For example, the review by Ouwens *et al.* reported on interventions or

programmes that aimed at improving care for adult patients with cancer. Patient-centeredness, an outcome not of interest to the overview, was presented as the proportion of studies with a significant difference between intervention and control group. The results were not synthesized quantitatively or narratively and in turn, the process of assessing the quality of evidence would be difficult (i.e., to assess the imprecision and inconsistency from a meta-analysis presented as a forest plot).

Lastly, the review had to report on a formal assessment of risk of bias (or methodological quality) of the included studies. This was because it is important that the reviews were performed with rigour and high methodological quality. Question 3.4 of the ROBIS tool, a tool to assess risk of bias in systematic reviews [101], was used. We proposed that if the answer to the question was "yes" or "probably yes", we could be more certain that the review was of low risk of bias and of high quality.

In summary, our inclusion criteria consisted of the adopted inclusion criteria from the meta-review by Martínez-González *et al.* [91]:

1. The review was a systematic review that reported on adult patients with any non-communicable chronic condition. Non-communicable chronic conditions, also known as chronic diseases, were defined as those not being passed from person to person, of long duration and generally slow progression; and
2. The review did not just report on addiction and mental disorders.

Additional inclusion criteria that we developed were that reviews had to:

3. Report on at least one specific chronic disease condition (e.g., diabetes or asthma would be acceptable, frailty or multimorbidity would not);

4. Include at least one randomized controlled trial reporting on a multidisciplinary team approach, defined as composed by two or more health care providers (e.g., nurse and physical therapist);
5. Report in a quantitative (meta-analysis) or narrative summary on at least one of the following patient-important outcomes: mortality, missed days of school or work, emergency room visits, length of in-patient stay, quality of life, a measure of functional status appropriate for the disease (e.g., comparable to joint damage or disease for hemophilia), educational attainment, patient adherence, or patient knowledge; and
6. Report a formal assessment of risk of bias (or methodological quality) of the included studies.

Methodological challenge: The inclusion criteria applied by the meta-review was too broad for the purpose of the overview.

Solution: We introduced four additional inclusion criteria to retrieve relevant and likely low risk of bias reviews.

3.3.3.2 Data extraction

Following the screening stage, we performed data extraction to retrieve relevant data from the included reviews. Two investigators (DZ and AW) independently performed data extraction, and a third investigator (CHTY) independently adjudicated all discrepancies.

The diverse presentation of primary study characteristics and results in the reviews presented a methodological challenge during the data extraction process. For example, some reviews reported the composition of the integrated multidisciplinary care team while others did not. As another example, some studies reported the event rate and number of participants, but not the risk estimates. The initial extraction form was modelled to collect information directly from the reviews with no intention to retrieve data from the primary studies. Data to be collected from the reviews included the study designs, study setting (e.g., country; and primary, secondary, tertiary, or community-based care setting), participant characteristics (age, ethnicity, and sex), components of the intervention and comparison groups (e.g., health care providers, and specific interventions such as education, physical training and rehabilitation, home care), and details on the all the reported outcomes and analyses. However, it was realized that it was not necessary to abstract all the details from each review. Instead, the overview was only concerned about data from our outcomes of interest (those identified from the NHF-McMaster Guideline on Care Models for Hemophilia Management) and general information about the review.

Therefore, to overcome this challenge a standardized extraction form was developed to systematically abstract the required data. The form consisted of two main parts. The first part required details from the review, including the authors, title, year of publication, and the population and intervention as reported by the review authors. The second part required information on the outcomes of interest from the review and the primary studies that reported on these outcomes. First, the outcomes of interest and the pooled estimates (relative risks (RR), odds ratios (OR), hazard ratios (HR), mean

differences (MD) and standardized mean differences (SMD)) and associated measures of dispersion, or narrative summaries were recorded. Then, for each outcome the primary studies that contributed to the evidence for the outcome was recorded. Details of the primary studies were noted, including participant characteristics (mean age and number of participants), study designs and length of follow-up, descriptions of the interventions (including which health care providers were involved in the delivery of care) and controls, and the risk of bias assessments of the studies. If the required information was not presented in the review, the extractor retrieved the data from the primary studies.

Methodological challenge: Due to the diverse presentation of data from the included reviews, the data extraction process would be difficult to retrieve relevant information for the overview.

Solution: We developed an extraction form that could collect relevant data that could be reported in different ways by the reviews. The form was divided into two main parts: the first to gather details on the review and the second for information on the outcomes of interest and the primary studies that reported on these outcomes.

3.3.4 Assessment of risk of bias of the included reviews

One of the factors judging whether a systematic review has low risk of bias is whether they formally assessed the risk of bias on their included studies. We had already assessed this factor at the screening stage of the overview, where we only included reviews that adequately applied a risk of bias tool. We were faced with a methodological challenge on whether to perform further risk of bias assessments on the reviews in other

domains. According to the ROBIS tool, there are four domains to consider and a concluding section to judge the overall risk of bias in the review. The four domains are as follows: to assess study eligibility criteria (domain 1), identification and selection of studies (domain 2), data collection and study appraisal (domain 3), and synthesis and findings (domain 4). The overall risk of bias includes a judgement of whether conclusions of the review were supported by the evidence which also considers the authors' interpretation of findings. We decided that risk of bias judgements in all domains were necessary to assess whether the evidence we were presenting were from low risk of bias reviews. We would have more confidence in the results reported by a low risk of bias review with rigorous methodology. For example, if their eligibility criteria was unambiguous, the search included an appropriate range of databases/electronic sources for published and unpublished reports, and additional methods to database searching was used to identify relevant reports, we would have more certainty in their published results.

At first, we planned to assess all four domains and the overall risk of bias. However, it was realized that the assessment of synthesis and findings of evidence (domain 4) was not necessary. All of the question prompts from domain 4 were already addressed by our overview. Question 4.1, "Did the synthesis include all studies that it should?" and question 4.3, "Was the synthesis appropriate given the nature and similarity in the research questions, study designs, and outcome across included studies?", would already be addressed in the overview. This was performed by returning to the primary studies and double-checked or performed a re-analysis if there was uncertainty about the syntheses presented by the review. Question 4.4, "Was between-study variation

(heterogeneity) minimal or addressed in the synthesis?" was also addressed by GRADE in the assessment of inconsistency. Question 4.6, "Were biases in primary studies minimal or addressed in the synthesis" was judged by the GRADE assessment of risk of bias. Additionally, overall risk of bias in the review was also declared unnecessary for assessment. All interpretations of findings and reporting of results were performed using GRADE. Therefore, it was not necessary to assess the review's interpretation of findings, the relevance of identified studies to their research question, and whether they avoided emphasizing results on the basis of statistical significance.

Overall, only domains 1, 2, and 3 were assessed using the ROBIS tool. We did not assess synthesis and findings of evidence (domain 4), or overall risk of bias (interpretation of findings) since these domains were considered when the overall quality of the evidence was assessed using GRADE.

It should also be noted that although we assessed domains 1, 2, and 3 using their ROBIS tool question prompts, not all questions were necessary to be answered. For example, question 3.2, "Were sufficient study characteristics available for both review authors and readers to be able to interpret the results?" and question 3.3, "Were all relevant study results collected for use in the synthesis?", were not important to be answered for our overview since we collected this information from the original studies if they were not adequately provided by the review.

For the risk of bias of the included reviews assessment, two investigators (CHTY, DZ) independently assessed the risk of bias of the reviews using the ROBIS tool [2], and a third investigator (NS) adjudicated all discrepancies.

Methodological challenge: In an overview of reviews, a quality assessment or risk of bias assessment of reviews is usually reported. However, we contemplated the necessity of completing such an assessment and whether to use all risk of bias factors to make the assessment.

Solution: We decided to perform a risk of bias assessment on the reviews since they provided some degree of certainty of how much we trust the reported results of the reviews. However, we did not assess all risk of bias factors using the ROBIS tool. Only domains 1, 2, and 3 were assessed using the ROBIS tool. Domain 4 and the overall risk of bias were assessed with GRADE when considering the quality of evidence.

3.3.5 Assessment of quality of evidence of each outcome

The next step of the overview process was to assess the quality of evidence of each outcome. The GRADE approach was applied with two investigators (CHTY and DZ), and a third investigator resolved any discrepancies (NS). The quality of evidence was assessed as high, moderate, low, or very low. The effect estimates or narrative summaries, and quality of evidence were summarized in the GRADE summary of findings table by outcome and by chronic condition. The evidence for each outcome was assessed using the GRADE domains: risk of bias, inconsistency and imprecision of the effect, indirectness and publication bias.

Risk of bias

Randomized controlled trials were the most common type of study that were reported in the included reviews. Few of the studies were non-randomized controlled trials or controlled before-after studies. We used the risk of bias assessments as performed and reported by the included reviews. Risk of bias assessments by the reviews was commonly performed using the Cochrane risk of bias tool and presented the judgements on sequence generation (selection bias), allocation sequence concealment (selection bias), blinding of participants and personnel (performance bias), blinding of outcome assessment (detection bias), incomplete outcome data (attrition bias), selective outcome reporting (reporting bias), and other potential sources of bias. If authors only reported the overall risk of bias, we retrieved the primary studies to determine which domain(s) had serious concerns. For our risk of bias assessment in this overview, we were mainly concerned with risk of bias due to random sequence generation, allocation concealment, blinding of outcome assessment, incomplete outcome data, and selective reporting. We did not focus on risk of bias due to blinding of participants and personnel since it is very difficult to blind the patients and health care providers to whether they were or were not receiving or delivering the integrated multidisciplinary care intervention. Some reviews also noted the unnecessary assessment of this domain and used the Health Technology Assessment-Disease Management (HTA-DM) tool [102]. The HTA-DM was specifically designed to only address concerns relevant to the administration of disease management programs. Another risk of bias tools used by the reviews was the Jadad score [103]. However, as recommended by the ROBIS tool, we only included reviews that used the

Jadad score and assessed allocation concealment. This was important since allocation concealment was not included in the Jadad score.

We considered rating down the quality of evidence when there was substantial risk of bias in the studies contributing the greatest for each outcome and noted which domains were of concern in the summary of findings table.

Inconsistency

Inconsistency was judged using the forest plots as presented by the reviews with quantitative syntheses (meta-analyses). Judgements were made based on the similarity of point estimates, the extent of overlap of CIs, and statistical criteria [62]. We used the statistical test of I^2 as reported by the reviews to judge the proportion of the variation in point estimates due to among-study differences. *A priori* explanations for substantial heterogeneity as defined by a large variation in effect, non-overlapping CIs, and large I^2 values. Subgroup analyses by the study authors were considered to explain potential heterogeneity. For instance, one review reported considerable heterogeneity ($I^2 = 93\%$) for an asthma quality of life score and performed a subgroup analysis of studies with greater integrated disease management (i.e., three disease management interventions, as compared to less than three interventions) [104]. The review found that with more disease management interventions, there was an even smaller effect. Therefore, we rated down the quality of evidence for inconsistency. Therefore, we rated down the quality of evidence for inconsistency. In another example, a review in heart failure found that risk of mortality was reduced in those receiving clinical service interventions (integrated

multidisciplinary care) compared to usual care [105]. They performed a subgroup analysis to only include studies with proper allocation concealment and the I^2 decreased greatly. We decided to use the studies that performed proper allocation concealment and did not downgrade for inconsistency.

Indirectness

GRADE identifies four ways that evidence could be indirect, which are: differences in the population (applicability), differences in the interventions (applicability), differences in outcome measures (surrogate outcomes), and indirect comparisons [63]. However, a methodological challenge appeared for assessing indirectness for our overview. It can be difficult to make judgements for systematic reviews. Systematic reviews will typically clearly specify the population and interventions of interest in their eligibility criteria to ensure that only direct relevant studies will be eligible. For our overview, although we were strict on our populations of interest (adults with chronic conditions) and outcomes (only those of interest to hemophilia and most chronic conditions), we accepted reviews that included at least one study with a multidisciplinary care team (two or more health care providers). We used this criterion in order to capture as many relevant studies as possible for the overview and guideline. Therefore, we considered rating down the evidence for indirectness if some of the studies included in the review did not include an active multidisciplinary team (e.g., only nurse-led care). Studies with composite outcomes were also considered for downgrading due to indirectness. For example, a review on heart failure reported a

composite of emergency department visits and hospital admissions, rather than emergency department visits only, a potentially indirect measure of our patient-important outcome for emergency department visits [106].

Methodological challenge: Typically, indirectness assessments are limited in reviews. However, we found that there could be potential sources of indirectness related to our overview question.

Solution: We assessed indirectness based on the intervention, whether the intervention involved an active multidisciplinary team, and the reported outcome. Using a consistent approach to assess indirectness based on these criteria revealed any sources of indirectness that could downgrade the quality of evidence.

Imprecision

Imprecision was considered by an examination of the 95% CIs and whether there were a sufficient number of events and sample size. Additionally, the CIs were used to judge whether clinical action would differ if the upper versus the lower boundary of the CI represented the truth [61]. For example, for one review on persons with COPD, functional status was measured by the mean change in the 6-Minute Walking Test (6MWT) with integrated disease management [42]. The studies reported a mean difference of 43.86 meters and a 95% CI of 21.84 to 65.89 meters. The minimal clinically important difference of the 6MWT was 35 meters. Since the 95% CI included clinically important and not important effects of integrated multidisciplinary management, we considered rating down the quality of evidence.

Publication bias

Publication bias was considered in reviews with small studies. Studies with greater number of participants are less likely to remain unpublished or ignored because they give more precise estimates of treatment effect [64]. We also considered rating down the quality of evidence if the review reported an asymmetrical funnel plot or other measures of publication bias such as considering whether all relevant search databases were used, the number participants from the included studies, and whether there could be publication bias by study sponsors.

3.3.6 Assessment of quality of evidence for the guideline

As discussed, our second objective of this overview was to use some of the evidence to inform decisions of the guideline panel to issue recommendations for the NHF-McMaster Guideline on Care Models for Hemophilia Management. However, there was a methodological challenge in considering how the evidence from the overview were to be used for the guideline. We determined that the evidence assessment for risk of bias, inconsistency, imprecision, and publication bias, would not change between the overview and guideline. However, an additional step was required for the guideline to judge the indirectness of evidence to the hemophilia care setting. We overcame this challenge by using an indirectness appraisal form that included descriptions of the studies contributing to each outcome. This included a description of the chronic condition population, intervention, and outcomes from the relevant studies from each review. The panel

members, blinded to the results and evidence assessments in the remaining domains, completed the form before the guideline recommendations meeting to judge whether the evidence from each review were sufficiently direct to hemophilia [28]. For example, the reviews including study population of heart failure patients was judged as not sufficiently direct to hemophilia since the mean age of heart failure patients was much older than the general hemophilia population.

3.4 Results

The literature search identified 1,191 non-duplicate records. After title and abstract screening of the search result set, 269 articles were assessed for full-text review, including 30 identified through other sources. Of these 30 articles, 27 were from the meta-review by Martínez-González *et al.* [91]. Altogether, 262 articles were excluded, and seven [42, 104–109] included. A PRISMA diagram of the selection flow is provided in Figure 7.

Of the seven included systematic reviews, two reported on COPD [42, 108]; two reported on asthma [107, 109]; one reported on both asthma and COPD [104]; and two reported on heart failure [105, 106]. Characteristics of the included reviews are described in Table 9.

We presented the results for each outcome by chronic condition. Overall, the quality of evidence ranged from high to low. We reported the results in a summary of findings table (Table 10).

Mortality

Kruis *et al.* [42] described five RCTs with 1,235 patients of any age with COPD. After a mean of 12 months, the OR for mortality with the pooled comparison of integrated disease management versus non-integrated care was 0.85; 95% CI: 0.49 to 1.46. The quality of evidence was moderate due to some inconsistency of results across studies and CIs including the potential for either increased or reduced mortality (Table 10). Peytremann-Bridevaux *et al.* [108] reported similar results from nine RCTs and one non-RCT with 1,102 adult patients with COPD (OR 0.84; 95% CI: 0.54-1.40). The quality of evidence was also moderate due to wide CIs (Table 10). Overall, the evidence from both reviews in COPD showed a probable reduction in mortality.

Health Quality Ontario [106] conducted a review of eight RCTs including 2,787 adults with heart failure. The meta-analysis showed a reduction in mortality with specialized multidisciplinary community-based care compared to usual care after a mean of 12 months follow-up (RR 0.71; 95% CI: 0.56 to 0.91). The authors reported similar results after subgroup analyses when the intervention was delivered through a direct team model (clinic). This evidence was rated as moderate due to some studies with unclear allocation concealment, and some interventions not involving an active multidisciplinary team (Table 10). Results were very similar to another review in heart failure by Takeda *et al.* [105], which included eight RCTs with 1,784 adults with heart failure. The review found that the risk of mortality was reduced in those receiving clinical service interventions as compared to usual care after a range of 6 to 12 months follow-up (OR 0.60; 95% CI: 0.43 to 0.84). The moderate heterogeneity that was observed (I^2 = 37%)

was partly explained by the risk of bias from lack of or unclear allocation concealment. The quality of evidence was high (Table 10). Overall, the quality of evidence was high in heart failure showing a reduction in mortality.

Missed days of school or work

Baishnab and Karner [107] provided two studies for this outcome. One RCT including 191 adults with asthma found for organized asthma clinics against standard of care a MD for the number of days lost from work over six months of -0.57 days; 95% CI: -1.57 to 0.48. The quality of this evidence was low due to imprecision with few events and the CIs including more, fewer, or no days lost. Peytremann-Bridevaux *et al.* [109] found two RCTs in 477 adults with asthma. The two studies reported comparable missed days between the intervention and usual care within 12 months of follow-up. The quality of the evidence was also low as some of the studies had unclear concealment, no blinding, and inadequate randomization, and the effect was imprecise (Table 10). Overall, the evidence in asthma showed that integrated care may reduce the number of missed days of school or work.

Emergency room visits

Two reviews reported number of emergency room visits in people with COPD. Kruis *et al.* [42] included four RCTs with 1,161 patients of any age with follow-up from 3 to 12 months. Integrated disease management compared to usual care for the number of patients with at least one emergency department visits showed an OR of 0.64; 95% CI:

0.33 to 1.25. The quality of evidence was moderate due to inconsistency across studies (I^2 = 71%) and some studies with incomplete outcome reporting or selective reporting (Table 10). Lemmens *et al.* [104] found four RCTs with 462 adult patients followed for 2 to 9 months. The analysis showed that multiple interventions of integrated disease management compared to standard of care had a reduction in the number of emergency department visits per patient with a MD of -0.08 visits; 95% CI: -0.18 to 0.03. The quality of evidence was low due to downgrading for some studies with unclear concealment, no blinding, and inadequate randomization; and for indirectness, since the studies assessing this specific outcome did not include an active multidisciplinary team as intervention (Table 10). Overall, the evidence in COPD showed that integrated care probably has little to no effect on emergency room visits.

Two reviews reported emergency room visits in people with asthma. Baishnab and Karner [107] pooled two RCTs with 344 adults attending an asthma clinic or not for 6 to 9 months. They found that an OR of 1.03 (95% CI: 0.21 to 5.15) for the number of people with one or more event prompting emergency department visits compared to standard of care. The quality of evidence was low due to unclear allocation concealment and no blinding; some inconsistency across studies (I^2 = 63%); and imprecision with the effect including more, fewer or similar number of visits (Table 10). Peytremann-Bridevaux *et al.* [109] included five RCTs, one non-RCT, and one controlled before-after study, with mean of 6 months follow-up. The data were summarized narratively in the review: one RCT found a reduction in emergency department or unscheduled visits; and four RCTs and one non-RCT did not find an effect. One controlled before-after study found no

important reduction between groups. The quality of evidence was moderate due to some studies with incomplete outcome reporting, and unclear or ineffective allocation concealment; and the number of visits in the intervention and control groups differed greatly between studies (Table 10). Overall, the evidence in asthma showed that there may be little to no difference in emergency room visits for those with integrated care.

One review reported visits in people with heart failure. Health Quality Ontario [106] included one RCT with 151 adults followed for 12 months. The study found that specialized multidisciplinary community-based reduced the number of readmissions or visits per person compared to care by a primary care physician (HR 0.67; 95% CI: 0.47 to 0.96). The quality of evidence was low from unclear allocation concealment; few events; and indirectness due to the intervention not involving an active multidisciplinary team and the outcome being a composite of emergency department visits and hospital admissions (Table 10). Overall, the evidence in heart failure showed that integrated care reduces emergency room visits.

Length of in-patient stay

Kruis *et al.* [42] found six RCTs of 741 patients of any age with COPD reporting the difference in mean hospitalization days per patient per group (integrated disease management versus control). Patients treated with integrated disease management were on average discharged from hospital nearly four days earlier compared to control patients (MD -3.78 days; 95% CI: -5.9 to -1.67). The quality of evidence was high (Table 10).

Overall, the evidence in COPD showed that integrated care reduced length of in-patient stay.

Health Quality Ontario [106] reported on seven RCTs with adults with heart failure. Meta-analysis could not be conducted and results were reported narratively. Overall, persons receiving specialized multidisciplinary community-based care had shorter hospital stays, although one study showed longer hospital stays. The quality of evidence was low due to no or unclear allocation concealment, some studies not involving an active multidisciplinary team in the intervention, and wide variation in the mean number of days reported by studies (Table 10). Overall, the evidence in heart failure showed length of stay may be reduced.

Quality of life

All seven reviews reported quality of life. Two reviews included people with COPD and/or asthma and quality of life was measured with the St. George's Respiratory Questionnaire (SGRQ, 0 to 100; with 100 indicating worse quality of life). Kruis *et al.* [42] included 13 RCTs (1,425 patients of all ages) followed for 3 to 12 months. The change in quality of life was greater with integrated care (MD -3.71 points; 95% CI: -5.83 to -1.59), which did not reach, but nor did it exclude, the minimally important difference of 4 points [110]. The quality of evidence was high (Table 10). Lemmens *et al.* [104] found five RCTs of adults with COPD and one RCT of adults with asthma, with a total of 770 patients followed for 3 to 12 months. There was little to no difference in quality of life (MD -2.52 points; 95% CI: -5 to -0.05), however, a minimally important difference of

4 points was not excluded. The quality of evidence was high (Table 10). Overall, the evidence from both reviews in COPD and/or asthma showed that quality of life was improved, although it may not be a clinically important improvement.

Two reviews included people with asthma. Baishnab and Karner [107] included two RCTs followed for 4 to 9 months. The results could not be pooled because one study used the SGRQ, while the other used the Asthma Quality of Life Questionnaire (AQLQ, 0 to 7; with 0 indicating worse quality of life) and EuroQol 4-Dimensions Questionnaire (EQ4D). Quality of life favoured those attending asthma clinics in one study; but the other study reported no change. The quality of evidence was moderate due to unclear allocation and no blinding, and few participants and wide CIs (Table 10). Lemmens *et al.* [104] included three RCTs and one CBA study with a total of 841 adults. Multiple disease management interventions improved quality of life as measured by the AQLQ (MD 0.35; 95% CI: 0.21 to 0.50). The quality of evidence was assessed as moderate due to inconsistency across studies (I^2 = 93%), and a subgroup analysis of studies with greater integrated disease management showing smaller effects (MD 0.15: 95% CI: -0.03 to 0.33) (Table 10). Peytremann-Bridevaux *et al.* [109] found eight RCTs (1,627 adults) with follow-up from 2 to 12 months and reported a SMD of 0.22; 95% CI: 0.08 to 0.37. On the AQLQ scale, the results represented a mean difference of 0.31 (0.11 to 0.53), which did not reach, but did not exclude, a minimal clinically important difference (0.5). The quality of evidence was moderate due to some studies with incomplete outcome data, unclear blinding, selective reporting and inadequate randomization (Table 10). The evidence in asthma showed that quality of life is probably improved.

Two reviews included adults with heart failure. Health Quality Ontario *et al.* [106] included five RCTs using the Minnesota Living with Heart Failure Questionnaire (MLHFQ) but results were not pooled and were summarised narratively. The studies showed that specialized multidisciplinary community-based care may improve quality of life, and may be seen in the physical domain, but not the emotional domain. The quality of evidence was assessed as low due to inconsistency between total mean scores and domain scores; and wide CIs for domain scores (Table 10). Takeda *et al.* [105] included nine RCTs and reported results narratively. Five of the studies reported very little difference in health related quality of life scores between the clinical service intervention and control groups. The quality of evidence was low due to some studies with incomplete outcome data, unclear blinding, selective reporting, and inadequate randomization; and some imprecision from the small proportion of patients completing the questionnaires at all time points (Table 10). In summary, the evidence in heart failure showed that integrated care may improve or have little to no difference in quality of life.

Functional status

Kruis *et al.* [42] included 14 RCTs with 871 patients of any age with COPD followed for 3 to 12 months. There was an improvement in functional exercise capacity as measured by the mean change in the 6MWT with integrated disease management (MD 43.86 meters; 95% CI: 21.83 to 65.89). The quality was moderate due to inconsistency across studies (I^2 = 83%) (Table 10). Removing low quality studies from the analysis reduced the difference in distance walked to 15.15 meters (6.37 to 23.93), which does not

reach minimal clinically important difference of 35 meters. Peytremann-Bridevaux *et al.* [108] included five RCTs with 359 adults with COPD followed a mean of 12 months. The difference with multidisciplinary disease management was 32.2 meters (95% CI: 4.1 to 60.3). The quality was moderate due to some studies with no or unclear allocation concealment and few people (Table 10). Overall, the evidence in COPD showed that functional status is probably improved.

Other outcomes: educational attainment, patient adherence, patient knowledge

We did not find reviews reporting educational attainment, patient adherence, and patient knowledge (Table 10).

Use of evidence in the NHF-McMaster Guideline on Care Models for Hemophilia Management

Some of the evidence included in the overview was used to inform decisions of the guideline panel issuing recommendations for the NHF-McMaster Guideline on Care Models for Hemophilia Management. The additional evidence appraisal step was a judgement on indirectness of evidence to the haemophilia care setting. An indirectness form was completed by the panel members before the panel meeting to judge whether the evidence from each review were sufficiently direct to hemophilia based on the population, intervention, and outcome. The evidence was in general sufficiently direct from reviews of people with asthma or COPD, but not heart failure (e.g., older population than persons with hemophilia, and different natural course of the disease); and some

evidence was not sufficiently direct due to the intervention (e.g., an integrated care team consisting only of a nurse and physician) or outcome (e.g., a measure of pain, which is important to hemophilia patients, was not included in the SGRQ quality of life questionnaire). Table 11 reported the GRADE evidence profiles used for the NHF-McMaster Guideline on Care Models for Hemophilia Management. Overall, when used for hemophilia, the evidence was of moderate quality for all the outcomes, with the exception of missed days of school or work, for which it was of very low quality (Table 11).

3.5 Conclusion

Overall, we analyzed seven systematic reviews in three chronic disease areas. We found high to low quality evidence for different patient important outcomes. We found that there was a reduction in mortality, and likely a reduction in emergency room visits and improvement in function with integrated care. There was little to no difference in quality of life, but shorter hospital stays, and there may be little to no difference in missed days of school or work. No reviews reported on educational attainment, patient adherence, nor patient knowledge. There was no high quality evidence for a negative direction of the effect (i.e., causing harm) for any of the outcomes or interventions related to the considered chronic conditions.

The strengths of this overview of reviews include the identification and inclusion of systematic reviews that were at low risk of bias and the use of GRADE to assess the level of evidence for each outcome from each review. As a result, readers can focus on

the best available evidence on important outcomes of integrated care. However, there are potential limitations of this review. First, we sought systematic reviews on integrated multidisciplinary care for adult patients with non-communicable chronic conditions, excluding populations with addiction and mental disorders. We were unable to include reviews on other chronic conditions with well-established integrated multidisciplinary care models such as stroke, due to adoption of a strict inclusion and exclusion criteria. The most common reasons for the exclusion were that they did not formally assess risk of bias (or methodological quality) of their primary studies or they did not report on one of our identified patient-important outcomes. Other populations with well-established integrated multidisciplinary care could be of interest to other stakeholders but are beyond the scope of our overview. These included populations with depression, HIV/AIDS, liver diseases, and other chronic conditions. Since this overview was limited to report on the evidence from the included reviews, we were unable to examine population subgroups that may have been of interest. This includes chronic condition populations of differing severity.

Second, we focused on integrated multidisciplinary care as the intervention of interest. We chose to focus on the meso-level dimension of integration proposed by Valentijn *et al.* [95], as previously discussed. The systematic reviews we included consisted of diverse interventions of integrated multidisciplinary care. It is likely that the composition of each intervention, such as the multidisciplinary health care team, presence of home care, initiation of a self-management and education program, are important to the

effects of integrated multidisciplinary care. However, our review does not address these individual components of the intervention.

Third, the outcomes of interest for this overview of reviews were chosen based on the NHF-McMaster Guideline on Care Models for Hemophilia Management development process. These outcomes were selected through a rigorous process of asking experts and patients in the field of hemophilia, and the same outcome categories are likely to be important to patients with other chronic conditions. Of course, functional outcomes need to be disease-specific. For example, the review by Peytremann-Bridevaux *et al.* [109] reported several asthma-specific outcomes, such as the asthma severity score and the number of asthma exacerbations. The review by Kruis *et al*. [42] also examined the lung function of COPD patients using the forced expiratory volume in 1 second (FEV1) and forced vital capacity (FVC) scores. Indeed, these outcomes can be seen as patient-important by their respective populations, and should be appraised by interested groups.

As with the systematic review in care models for hemophilia management, we overcame various methodological challenges with the development of this overview of reviews. Many of these challenges were due to adopting review and GRADE methodology to the relatively novel concept an overview. Additional challenges were presented as a result of the unique process of using this evidence for hemophilia. Overall, we have provided an example of how indirect evidence could be used to inform the recommendations of a guideline in rare diseases.

Chapter 4. Methodological challenges and facilitators in the development of a guideline in rare diseases and a discussion of additional strategies

4.1 Introduction

The objective of this thesis was to identify the methodological challenges and facilitators, and decision-making process in determining solutions for the guideline development of care models for hemophilia management. Chapters 2 and 3 discuss the methodological challenges and solutions that were applied in the evidence gathering and assessment stages of guideline development. The aim of this thesis chapter is to summarize these challenges and solutions in the context of the facilitators and barriers to clinical care, research, and guideline development in rare diseases as introduced in chapter 1 (Table 1).

4.2 Summary of the methodological challenges and facilitators

In chapter 1, overall facilitators and barriers to clinical care, research, and guideline development in rare diseases were identified (Table 1). Many of these barriers were presented in the evidence gathering and assessment stages of the guideline in care models for hemophilia management. We applied solutions and presented our approach to demonstrate how barriers in the development of a guideline in rare diseases could be overcome. On a broader scale, the methodological challenges we faced could be categorized into the common barriers in clinical care, research, and guideline development for rare diseases.

4.2.1 Barriers in clinical care, research, and guideline development

Many of the challenges that were faced in the hemophilia systematic review were due to the paucity of evidence on care models. One of the main reasons for the lack of evidence was that policymakers, funders, and researchers in both the clinical care and research phase typically do not set a priority on studying care models in hemophilia (Table 1). Integrated care has been established and widely accepted since the 1970s and few question its efficacy. However, it is possible that one care model over another may be more beneficial, especially for certain subgroups, and there may be potential harm associated with other care models. There may also be additional resource, cost, and equity concerns that may be associated with the care models. Another reason why care models in hemophilia are seldom studied is that rare diseases tend to be lower priority in research and initiatives in the public health sector as compared to more common conditions. Orphan drugs, or products, are examples of consequences of neglected rare disease studies. Orphan products were termed by the notion that rare diseases in small patient populations are "orphaned" by the pharmaceutical industry due to the large cost of drug development and low return on investment [111]. Pharmaceutical companies view a better return on investment with products for more common conditions, as they are more likely to reach the marketplace and are at greater demand as compared to those for rare diseases [112]. As a result, few products are developed and available for persons with rare diseases. There is a lack of research investment in hemophilia as a rare disease and care models in hemophilia are viewed as a low priority for policymakers, funders and researchers. Therefore, the three methodological challenges in the systematic review on

care models in hemophilia management arose. First, designing a search strategy with comprehensive non-integrated care model key terms was difficult, since care models in hemophilia were not well-defined in the literature as they were not a priority to study. Second, the methodological challenge of retrieving as many studies on care models in hemophilia was due to the notion that there are limited studies focusing on this research question. Third, authors tended to publish the results of their studies in care models in hemophilia in multiple reports during the 1970s and 1980s, but are seldom conducted and published today as they are not viewed as a priority.

Another barrier that affected both the clinical care and research phase was the perceived lack of clinical equipoise when conducting studies in rare diseases (Table 1). Since integrated care is widely accepted, it would be seen as unethical to propose a high quality randomized controlled trial with the other care models. As a result, there are many studies in integrated care, but not many in non-integrated care. This led to further methodological challenges for the systematic review on care models in hemophilia management. First, there was difficulty in retrieving non-integrated care studies. Second, we had used terms to describe integrated care for the search strategy that could have introduced bias, but were necessary to capture the widely reported care model.

Common in the study of rare diseases is the inherent heterogeneity in studies (Table 1). Heterogeneity could be due to the few studies in rare diseases, therefore the likelihood of the few studies reporting their outcomes differently increases. It would be more difficult to pool the results with different outcome measures. For more common conditions, even if there are various outcome measures, there could be enough studies

that assess each outcome measure and an overall estimate could be provided for each. Without standardized approaches to measure some patient important outcomes, we were presented with the methodological challenge to extract, summarize, and decisions on how to use the data from studies with different objectives to study care models in hemophilia. Additionally, since the outcomes of interest for the systematic review may have been embedded elsewhere in the text, for instance in the discussion, screeners must be trained thoroughly to look for these outcomes.

The paucity of published data on hemophilia and care models, which was also low quality evidence, affected the guideline development phase (Table 1). With the evidence limited to non-randomized controlled trials, we decided to use evidence from other chronic conditions. The lack of evidence for critically important outcomes in hemophilia, quality of life, educational attainment, and patient adherence were thought to be reported for other chronic conditions. This led to the development of an overview of reviews in integrated multidisciplinary care for the management of chronic conditions. The overview of reviews also had many methodological challenges including: deciding the scope of the overview, whether to use the search strategy and results of a previously conducted meta-review, clarifying the unit of analysis, defining strict inclusion criteria to identify relevant and manageable reviews for synthesis, developing a unique standardized extraction form for overviews of reviews, decisions on how to conduct an assessment of risk of bias of the included reviews, and how to assess indirectness for our overview question.

4.2.4 Overall facilitators

Although there were many barriers in the evidence gathering and assessment stages for the guideline, the process seemingly benefitted from hemophilia being a rare condition. One typical barrier in the research stage for rare diseases is the difficulty of study enrolment where patients are not registered in database in a reliable, harmonized way (Table 1). However, the recording of patient characteristics and data is well-established in HTCs, which is unique to hemophilia and some other rare diseases. For this reason, national studies such as the study by Soucie *et al.* [54] used data from registries recorded by 130 HTCs in the US. Although those attending non-integrated care were not as well recorded, the study authors were still able to retrieve data from physicians' offices, laboratories, pharmacies, hospitals, emergency rooms, and outpatient clinics, recruiting approximately 974 patients, 33% of the total recruited hemophilia population. In some cases of rare diseases and as was with hemophilia, patients are highly engaged with their treatment and ongoing research [113]. Hemophilia patients in integrated care especially, could be more willing to participate in registries.

Another facilitator throughout the guideline development process was having dedicated hemophilia experts from the Core Methods Group and guideline panel as a source of valuable information. Although a barrier in clinical care and research is the lack of clinical expertise in rare diseases, we were able to acquire experts in the field that provided insight into non-integrated care, studies that we may have missed, and indirectness assessments with other chronic conditions to the hemophilia setting.

4.3 Conclusion to thesis

This thesis has identified the methodological challenges in the development of a guideline in a rare disease, such as hemophilia. Solutions that were applied were presented in order to provide strategies for further guideline development in rare diseases and to transparently show our decision-making process. Through this process, it was demonstrated that it is possible to successfully gather and synthesize evidence for a rare disease with a paucity of high quality studies.

It was briefly discussed that additional strategies were applied to the evidence gathering and assessment stages of this guideline. These included using sufficiently direct evidence from other chronic conditions, systematic observation forms and expert-based evidence, and qualitative interviews to inform the guideline. Further innovative strategies to gather and synthesize evidence for guidelines in rare diseases can be developed, such as a greater use of registry data, and specifically developing patient values and preferences searches with panel assessments. As more strategies are developed, it is anticipated that more methodological challenges to incorporate this data will arise and add to the complexity of guideline development in rare diseases. However, the end result will be that high quality guidelines can be developed for rare diseases.

References

1. Yeung CHT, Santesso N, Pai M, Kessler C, Key NS, Makris M, et al. Care models in the management of haemophilia: a systematic review. Haemophilia. 2016;22:31–40.

2. Yeung CHT, Santesso N, Zeraatkar D, Wang A, Pai M, Sholzberg M, et al. Integrated multidisciplinary care for the management of chronic conditions in adults: an overview of reviews and an example of using indirect evidence to inform clinical practice recommendations in the field of rare diseases. Haemophilia. 2016;22:41–50.

3. Aymé S, Rodwell C, editors. 2013 Report on the State of the Art of Rare Disease Activities in Europe [Internet]. European Union Committee of Experts on Rare Diseases (EUCERD); 2013. Available from: http://www.eucerd.eu/upload/file/Reports/2013ReportStateofArtRDActivities.pdf

4. Global Genes: Allies in Rare Disease. RARE List [Internet]. Available from: https://globalgenes.org/rarelist/

5. National Center for Advancing Translational Sciences. Genetic and Rare Disease Information Center (GARD) [Internet]. Available from: https://rarediseases.info.nih.gov/gard/categories

6. Orphan Drug Act [Internet]. U.S. Food and Drug Administration; 2013. Available from: http://www.fda.gov/ForIndustry/DevelopingProductsforRareDiseasesConditions/HowtoapplyforOrphanProductDesignation/ucm364750.htm

7. The European Parliament and the Council of the European Union. Regulation (EC) No 141/2000 of the European Parliament and of the Council of 16 December 1999 on orphan medicinal products. OJ C. 2000;18:1–5.

8. Angelis A, Tordrup D, Kanavos P. Socio-economic burden of rare diseases: A systematic review of cost of illness evidence. Health Policy. 2015;119(7):964–79.

9. Baumann U, Stocklossa C, Greiner W, von der Schulenburg J-MG, von der Hardt H. Cost of care and clinical condition in paediatric cystic fibrosis patients. J Cyst Fibros. 2003;2(2):84–90.

10. Wildhagen MF, Verheij JB, Verzijl JG, Hilderink HB, Kooij L, Tijmstra T, et al. Cost of care of patients with cystic fibrosis in The Netherlands in 1990-1. Thorax. 1996;51(3):298–301.

11. Ginsberg G, Blau H, Kerem E, Springer C, Kerem BS, Akstein E, et al. Cost-benefit analysis of a national screening programme for cystic fibrosis in an Israeli population. Health Econ. 1994;3(1):5–23.

12. Baltin CT, Smaczny C, Wagner TO. [Drug treatment of cystic fibrosis - cost patterns and savings potential for outpatient treatment]. Med Klin (Munich). 2010;105(12):887–900.

13. Tur-Kaspa I, Aljadeff G, Rechitsky S, Grotjan HE, Verlinsky Y. PGD for all cystic fibrosis carrier couples: novel strategy for preventive medicine and cost analysis. Reprod Biomed Online. 2010;21(2):186–95.

14. Pai M, Iorio A, Meerpohl J, Taruscio D, Laricchiuta P, Mincarone P, et al. Developing methodology for the creation of clinical practice guidelines for rare diseases: A report from RARE-Bestpractices. Rare Diseases. 2015;3(1):e1058463.

15. Taruscio D, Morciano C, Laricchiuta P, Mincarone P, Palazzo F, Leo C, et al. RARE-Bestpractices: a platform for sharing best practices for the management of rare diseases. Orphanet J Rare Dis. 2014;9(Suppl 1):O14.

16. Sejersen T, Giovane CD, Filippini G, Leo CG, Meerpohl JJ, Mincarone P, et al. Methodology for production of best practice guidelines for rare diseases. Rare Diseases and Orphan Drugs. 2014;1(1).

17. The Lewin Group, Inc. National Hemophilia Foundation: Strategic Summit Report [Internet]. National Hemophilia Foundation; 2012 [cited 2016 Apr 11]. Available from: https://www.hemophilia.org/sites/default/files/article/documents/HemophiliaSummitFinalReportOct2012.pdf

18. Merritt A, Conneally P, Hilgartner M. Hemophilia: Genetics and Counseling. In: Hemophilia in Children. Littleton, MA: Publishing Sciences Group; 1976.

19. Rosendaal FR, Briët E. The increasing prevalence of haemophilia. Thromb Haemost. 1990;63(1):145.

20. Di Minno MND, Ambrosino P, Franchini M, Coppola A, Di Minno G. Arthropathy in patients with moderate hemophilia a: a systematic review of the literature. Semin Thromb Hemost. 2013;39(7):723–31.

21. Forsyth AL, Gregory M, Nugent D, Garrido C, Pilgaard T, Cooper DL, et al. Haemophilia Experiences, Results and Opportunities (HERO) Study: survey methodology and population demographics. Haemophilia. 2014;20(1):44–51.

22. Cassis FRMY, Querol F, Forsyth A, Iorio A, HERO International Advisory Board. Psychosocial aspects of haemophilia: a systematic review of methodologies and findings. Haemophilia. 2012;18(3):e101–14.

23. Hoots WK. Comprehensive care for hemophilia and related inherited bleeding disorders: why it matters. Curr Hematol Rep. 2003;2(5):395–401.

24. Bonser GM, Biggs RP, Willis R. William Goldie, 31 December 1907--13 February 1970. J Pathol. 1971;105(1):69–74.

25. Evans D. Haemophilia centres. BMJ. 285:1420.

26. Biggs R. Haemophilia treatment in the United Kingdom from 1969 to 1974. Br J Haematol. 1977;35(4):487–504.

27. Pai M, Key NS, Skinner M, Curtis R, Feinstein M, Kessler C, et al. NHF-McMaster guideline on care models for haemophilia management. Haemophilia. 2016;22(S3):6–16.

28. Pai M, Santesso N, Yeung CHT, Lane S, Schünemann HJ, Iorio A. Methodology for the development of the NHF-McMaster guideline on care models for haemophilia management. Haemophilia. 2016;22(S3):17–22.

29. Brozek JL, Akl EA, Alonso-Coello P, Lang D, Jaeschke R, Williams JW, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines. Part 1 of 3. An overview of the GRADE approach and grading quality of evidence about interventions. Allergy. 2009;64(5):669–77.

30. Brozek JL, Akl EA, Jaeschke R, Lang DM, Bossuyt P, Glasziou P, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines: Part 2 of 3. The GRADE approach to grading quality of evidence about diagnostic tests and strategies. Allergy. 2009;64(8):1109–16.

31. Brożek JL, Akl EA, Compalati E, Kreis J, Terracciano L, Fiocchi A, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines part 3 of 3. The GRADE approach to developing recommendations. Allergy. 2011;66(5):588–95.

32. Lane S, Sholapur N, Pai M, Iorio A. Outcomes, equity, acceptability, feasibility: a qualitative approach to identify and undertand stakeholder perceptions and experiences of models of care for hemophilia management in the U.S. [in progress].

33. Canadian Hemophilia Standards Group. Canadian Comprehensive Care Standards for Hemophilia and Other Inherited Bleeding Disorders, First Edition [Internet].

2007 [cited 2016 Apr 14]. Available from: http://www.hemophilia.ca/files/ComprehensiveCareStandards-EN.pdf

34. Street A. Developing models of haemophilia care. Haemophilia. 2012;18 Suppl 4:89–93.

35. Colvin BT, Astermark J, Fischer K, Gringeri A, Lassila R, Schramm W, et al. European principles of haemophilia care. Haemophilia. 2008;14(2):361–74.

36. Evatt BL, Black C, Batorova A, Street A, Srivastava A. Comprehensive care for haemophilia around the world. Haemophilia. 2004;10 Suppl 4:9–13.

37. De Kleijn P, Odent T, Berntorp E, Hilliard P, Pasta G, Srivastava A, et al. Differences between developed and developing countries in paediatric care in haemophilia. Haemophilia. 2012;18:94–100.

38. Geraghty S, Dunkley T, Harrington C, Lindvall K, Maahs J, Sek J. Practice patterns in haemophilia A therapy – global progress towards optimal care. Haemophilia. 2006;12(1):75–81.

39. Santiago-Borrero, Ortiz, Rivera-Caragol, Maldonado. Financial aspects of haemophilia care in Puerto Rico and other Latin American countries. Haemophilia. 1999;5(6):386–91.

40. Sooksriwong C, Boonkerd L, Chanjaruporn F, Chuansumrit A. Incremental cost effectiveness analysis for Haemophilia Home-Based Care Programme in Thailand. Haemophilia. 2012;18(4):e362–3.

41. Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011] [Internet]. The Cochrane Collaboration, 2011; [cited 2016 Apr 14]. Available from: www.cochrane-handbook.org

42. Kruis AL, Smidt N, Assendelft WJJ, Gussekloo J, Boland MRS, Rutten-van Mölken M, et al. Integrated disease management interventions for patients with chronic obstructive pulmonary disease. Cochrane Database Syst Rev. 2013;10:CD009437.

43. Arnold E, Lane S, Webert KE, Chan A, Walker I, Tufts J, et al. What should men living with haemophilia need to know? The perspectives of Canadian men with haemophilia. Haemophilia. 2014;20(2):219–25.

44. Drake JH, Soucie JM, Cutter SC, Forsberg AD, Baker JR, Riske B. High School Completion Rates Among Men with Hemophilia. American Journal of Preventive Medicine. 2010;38(4):S489–94.

45. Monahan PE, Baker JR, Riske B, Soucie JM. Physical functioning in boys with hemophilia in the U.S. Am J Prev Med. 2011;41(6 Suppl 4):S360–8.

46. Poon J-L, Zhou Z-Y, Doctor JN, Wu J, Ullman MM, Ross C, et al. Quality of life in haemophilia A: Hemophilia Utilization Group Study Va (HUGS-Va). Haemophilia. 2012;18(5):699–707.

47. Niu X, Poon JL, Riske B, Zhou ZY, Ullman M, Lou M, et al. Physical activity and health outcomes in persons with haemophilia B. Haemophilia. 2014;20(6):814–21.

48. Zhou Z-Y, Koerper MA, Johnson KA, Riske B, Baker JR, Ullman M, et al. Burden of illness: direct and indirect costs among persons with hemophilia A in the United States. Journal of Medical Economics. 2015;18(6):457–65.

49. Panicucci F, Baicchi U, Sagripanti A. Three years' experience in haemophilia home treatment. Scandinavian Journal of Haematology. 1977;19(S30):81–4.

50. Panicucci F. Hemophilia Home Treatment in Italy. Scandinavian Journal of Haematology. 1977;19(S31):25–7.

51. Smith PS, Keyes NC, Forman EN. Socioeconomic evaluation of a state-funded comprehensive hemophilia-care program. N Engl J Med. 1982;306(10):575–9.

52. Smith PS, Levine PH. The benefits of comprehensive care of hemophilia: a five-year study of outcomes. Am J Public Health. 1984;74(6):616–7.

53. Lazerson J. Hemophilia home transfusion program: effect on school attendance. J Pediatr. 1972;81(2):330–2.

54. Soucie JM, Nuss R, Evatt B, Abdelhak A, Cowan L, Hill H, et al. Mortality among males with hemophilia: relations with source of medical care. The Hemophilia Surveillance System Project Investigators. Blood. 2000;96(2):437–42.

55. Levine PH, Britten AFH. Supervised patient-management of hemophilia: a study of 45 patients with hemophilia A and B. Ann Intern Med. 1973;78(2):195–201.

56. Weiss HM, Simon R, Levi J, Forster Al, Hubbard M, Aledort L. Compliance in a comprehensive hemophilia center and its implications for home care. Family Systems Medicine. 1991;9(2):111–20.

57. Isarangkura P, Chuansumrit A, Panthangkura W, Hathirat P, Pandhawong S. Home therapy for hemophilia in Thailand. Southeast Asian J Trop Med Public Health. 1987;18(4):552–7.

58. Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. Journal of clinical epidemiology. 2011;64(4):401–6.

59. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 2011;64(4):383–94.

60. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). Journal of Clinical Epidemiology. 2011;64(4):407–15.

61. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6. Rating the quality of evidence—imprecision. Journal of Clinical Epidemiology. 2011;64(12):1283–93.

62. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 7. Rating the quality of evidence—inconsistency. Journal of Clinical Epidemiology. 2011;64(12):1294–302.

63. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 8. Rating the quality of evidence—indirectness. Journal of Clinical Epidemiology. 2011;64(12):1303–10.

64. Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. GRADE guidelines: 5. Rating the quality of evidence—publication bias. Journal of Clinical Epidemiology. 2011;64(12):1277–82.

65. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, et al. Evaluating non-randomised intervention studies. Health Technol Assess. 2003;7(27):iii – x, 1–173.

66. Sanderson S, Tatt ID, Higgins JPT. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. Int J Epidemiol. 2007;36(3):666–76.

67. Proposed Evaluation Tools for COMPUS: assessement, November 29, 2005. Canadian Coordinating Office for Health Technology, Ottawa, Canada; 2005.

68. Systems to rate the strength of scientific evidence [evidence report/technology assessment no 47]. AHRQ Publication No 02-E016. Agency for Healthcare Research and Quality, Rockville, MD; 2002;

69. Sterne J, Higgins J, Reeves B on behalf of the development groups for A-N. A Cochrane Risk of Bias Assessment Tool: for Non-Randomized Studies of

Interventions (ACROBAT-NRSI) [Internet]. Version 1.0.0, 24 September 2014; [cited 2016 Apr 18]. Available from: http://www.riskofbias.info

70. Sterne J, Higgins J, Reeves B on behalf of the development groups for R-I. The Risk Of Bias In Non-randomized Studies - of Interventions (ROBINS-I) assessment tool [Internet]. Version 7 March 2016; [cited 2016 Apr 18]. Available from: http://www.riskofbias.info

71. Soucie JM, Cianfrini C, Janco RL, Kulkarni R, Hambleton J, Evatt B, et al. Joint range-of-motion limitations among young males with hemophilia: prevalence and risk factors. Blood. 2004;103(7):2467–73.

72. Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, et al. GRADE guidelines: 9. Rating up the quality of evidence. Journal of Clinical Epidemiology. 2011;64(12):1311–6.

73. Mahlangu JN, For and on Behalf of the Medical and Scientific Council of the South African Haemophilia Foundation. Haemophilia care in South Africa: 2004–2007 look back. Haemophilia. 2009;15(1):135–41.

74. Chuansumrit A, Isarangkura P, Chantanakajornfung A, Panthangkool W, Hathirat P. Home treatment for patients with congenital bleeding disorders in a developing country. J Med Assoc Thai. 1999;82 Suppl 1:S57–62.

75. Kennelly JM, Tolley KH, Ghani AC, Sabin CA, Maynard AK, Lee CA. Hospital costs of treating haemophilic patients infected with HIV. AIDS. 1995;9(7):787–93.

76. Strawczynski H, Stachewitsch A, Morgenstern G, Shaw ME. Delivery of care to hemophilic children: home care versus hospitalization. Pediatrics. 1973;51(6):986–91.

77. Levine PH. The Home Therapy Program at the New England Area Hemophilia Center. Scandinavian Journal of Haematology. 1977;19(S31):37–51.

78. Levine PH. 8. The home therapy program at the New England area hemophilia center. Scand J Haematol Suppl. 1977;31:37–51.

79. Ekert H, Moorehead M, Williamson G. Home treatment of haemophilia. A follow-up study. Med J Aust. 1981;2(1):21–3.

80. Ekert H, Smibert E. “Home care” for haemophilia. Med J Aust. 1974;2(22):803–6.

81. Hilgartner MW. Home Care for Hemophilia: Current State of the Art. Scandinavian Journal of Haematology. 1977;19(S30):58–64.

82. Szucs TD, Offner A, Kroner B, Giangrande P, Berntorp E, Schramm W. Resource utilisation in haemophiliacs treated in Europe: results from the European Study on Socioeconomic Aspects of Haemophilia Care. The European Socioeconomic Study Group. Haemophilia. 1998;4(4):498–501.

83. Szucs TD, Öffner A, Schramm W. Socioeconomic impact of haemophilia care: results of a pilot study. Haemophilia. 1996;2(4):211–7.

84. Rabiner SF, Telfer MC, Fajardo R. Home transfusions of hemophiliacs. JAMA. 1972;221(8):885–7.

85. Rabiner SF, Telfer MC. Home transfusion for patients with hemophilia A. N Engl J Med. 1970;283(19):1011–5.

86. Soucie JM, Symons J, Evatt B, Brettler D, Huszti H, Linden J, et al. Home-based factor infusion therapy and hospitalization for bleeding complications among males with haemophilia. Haemophilia. 2001;7(2):198–206.

87. Tencer T, Roberson C, Duncan N, Johnson K, Shapiro A. A haemophilia treatment centre-administered disease management programme in patients with bleeding disorders. Haemophilia. 2007;13(5):480–8.

88. Abdullaev GM, Efendiev ZI, Askerova FR. [A five-year experience with the management of a group of patients with hemophilia in a special boarding school]. Gematol Transfuziol. 1983;28(6):49–51.

89. Carter F, Forbes CD, MacFarlane JD, Prentice CR. Cost of management of patients with haemophilia. Br Med J. 1976;2(6033):465–7.

90. Heemstra HE, Zwaan T, Hemels M, Feldman BM, Blanchette V, Kern M, et al. Cost of severe haemophilia in Toronto. Haemophilia. 2005;11(3):254–60.

91. Martínez-González NA, Berchtold P, Ullman K, Busato A, Egger M. Integrated care programmes for adults with chronic conditions: a meta-review. International Journal for Quality in Health Care. 2014;26(5):561–70.

92. McKenzie LL, Fie R, van Eys J. Medical, economic, and social impact of a home therapy program for hemophilia A on selected patients. South Med J. 1974;67(5):555–9.

93. Ingram GI, Dykes SR, Creese AL, Mellor P, Swan AV, Kaufert JK, et al. Home treatment in haemophilia: clinical, social and economic advantages. Clin Lab Haematol. 1979;1(1):13–27.

94. Strandberg-Larsen M, Krasnik A. Measurement of integrated healthcare delivery: a systematic review of methods and future research directions. Int J Integr Care

[Internet]. 2009 [cited 2016 Apr 23];9. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2663702/

95. Valentijn PP, Schepman SM, Opheij W, Bruijnzeels MA. Understanding integrated care: a comprehensive conceptual framework based on the integrative functions of primary care. International Journal of Integrated Care [Internet]. 2013 [cited 2016 Apr 18];13(1). Available from: http://www.ijic.org/article/10.5334/ijic.886/

96. Cohen E, Bruce-Barrett C, Kingsnorth S, Keilty K, Cooper A, Daub S. Integrated complex care model: lessons learned from inter-organizational partnership. Healthc Q. 2011;14 Spec No 3:64–70.

97. Baker JR. Design for an Evaluation of Hemophilia Treatment Center 340B Drug Pricing Programs. eScholarship [Internet]. 2013 [cited 2016 Apr 23]; Available from: http://escholarship.org/uc/item/01p6s37v

98. van Wetering CR, Hoogendoorn M, Mol SJM, Rutten-van Mölken MPMH, Schols AM. Short- and long-term efficacy of a community-based COPD management programme in less advanced COPD: a randomised controlled trial. Thorax. 2010;65(1):7–13.

99. Higginson IJ, Finlay I, Goodwin DM, Cook AM, Hood K, Edwards AGK, et al. Do hospital-based palliative teams improve care for patients or families at the end of life? J Pain Symptom Manage. 2002;23(2):96–106.

100. Boult C, Green AF, Boult LB, Pacala JT, Snyder C, Leff B. Successful Models of Comprehensive Care for Older Adults with Chronic Conditions: Evidence for the Institute of Medicine's "Retooling for an Aging America" Report. Journal of the American Geriatrics Society. 2009;57(12):2328–37.

101. Whiting P, Savovic J, Higgins JP, Caldwell DM, Reeves BC, Shea B, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. Journal of clinical epidemiology. 2016;69:225–34.

102. Steuten LMG, Vrijhoef HJM, van Merode GG, Severens JL, Spreeuwenberg C. The Health Technology Assessment-disease management instrument reliably measured methodologic quality of health technology assessments of disease management. Journal of Clinical Epidemiology. 2004;57(9):881–8.

103. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials. 1996;17(1):1–12.

104. Lemmens KMM, Nieboer AP, Huijsman R. A systematic review of integrated use of disease-management interventions in asthma and COPD. Respir Med. 2009;103(5):670–91.

105. Takeda A, Taylor SJC, Taylor RS, Khan F, Krum H, Underwood M. Clinical service organisation for heart failure. Cochrane Database Syst Rev. 2012;9:CD002752.

106. Health Quality Ontario. Community-based care for the specialized management of heart failure: an evidence-based analysis. Ontario Health Technology Assessment Series. 2009;9(17):1–42.

107. Baishnab E, Karner C. Primary care based clinics for asthma. Cochrane Database Syst Rev. 2012;4:CD003533.

108. Peytremann-Bridevaux I, Staeger P, Bridevaux P-O, Ghali WA, Burnand B. Effectiveness of chronic obstructive pulmonary disease-management programs: systematic review and meta-analysis. Am J Med. 2008;121(5):433–43.e4.

109. Peytremann-Bridevaux I, Arditi C, Gex G, Bridevaux P-O, Burnand B. Chronic disease management programmes for adults with asthma. Cochrane Database of Systematic Reviews [Internet]. 2015 [cited 2016 Apr 11]; Available from: http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD007988.pub2/abstract

110. Jones PW, Quirk FH, Baveystock CM. The St George's Respiratory Questionnaire. Respir Med. 1991;85 Suppl B:25–31; discussion 33–7.

111. Sharma A, Jacob A, Tandon M, Kumar D. Orphan drug: Development trends and strategies. J Pharm Bioallied Sci. 2010;2(4):290–9.

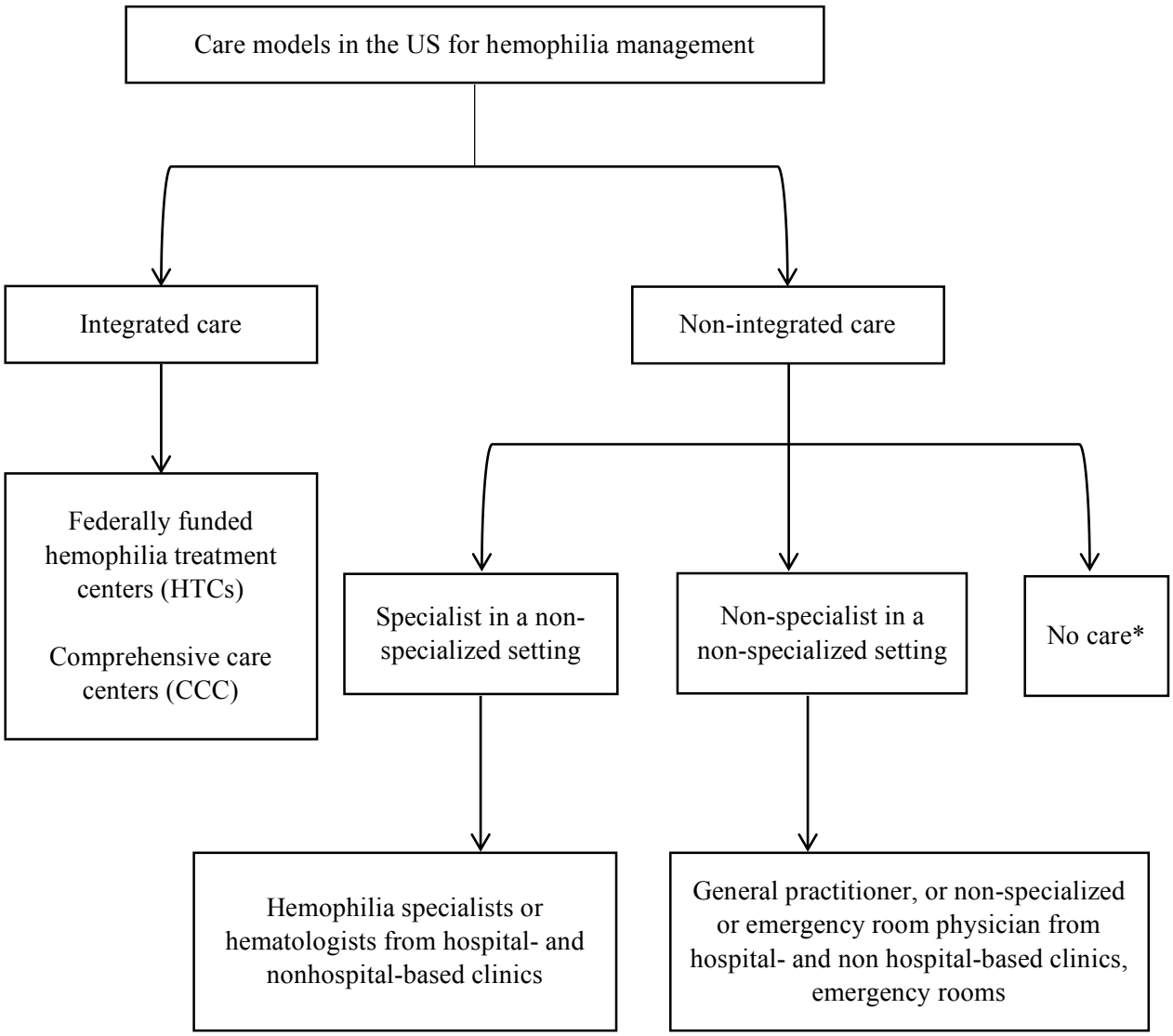
112. Groft S, de la Paz MP. Rare diseases - avoiding misperceptions and establishign realities: the need for reliable epidemiological data. In: Rare Diseases Epidemiology, Advances in Experimental Medicine and Biology. Springer Science and Business Media; 2010.

113. Gerß JWO, Köpcke W. Clinical Trials and Rare Diseases. In: Paz MP de la, Groft SC, editors. Rare Diseases Epidemiology [Internet]. Springer Netherlands; 2010 [cited 2016 May 2]. p. 173–90. Available from: http://link.springer.com.libaccess.lib.mcmaster.ca/chapter/10.1007/978-90-481-9485-8_11

114. Schünemann HJ, Wiercioch W, Etxeandia I, Falavigna M, Santesso N, Mustafa R, et al. Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. CMAJ. 2014;186(3):E123–42.

115. Pai M, Iorio A, Meerpohl J, Taruscio D, Laricchiuta P, Mincarone P, et al. Developing methodology for the creation of clinical practice guidelines for rare diseases: A report from RARE-Bestpractices. Rare Diseases. 2015;3(1):e1058463.

116. Martínez-Murillo C, Quintana S, Ambriz R, Benitez H, Berges A, Collazo J, et al. An economic model of haemophilia in Mexico. Haemophilia. 2004;10(1):9–17.



*The "no care" model represents complete absence of dedicated care, which does not appear to be operating in the Western World, but exists in other areas of the world where persons with hemophilia do not have access to care due to profound resource constraints

Figure 1. Observed care model organization in the United States.

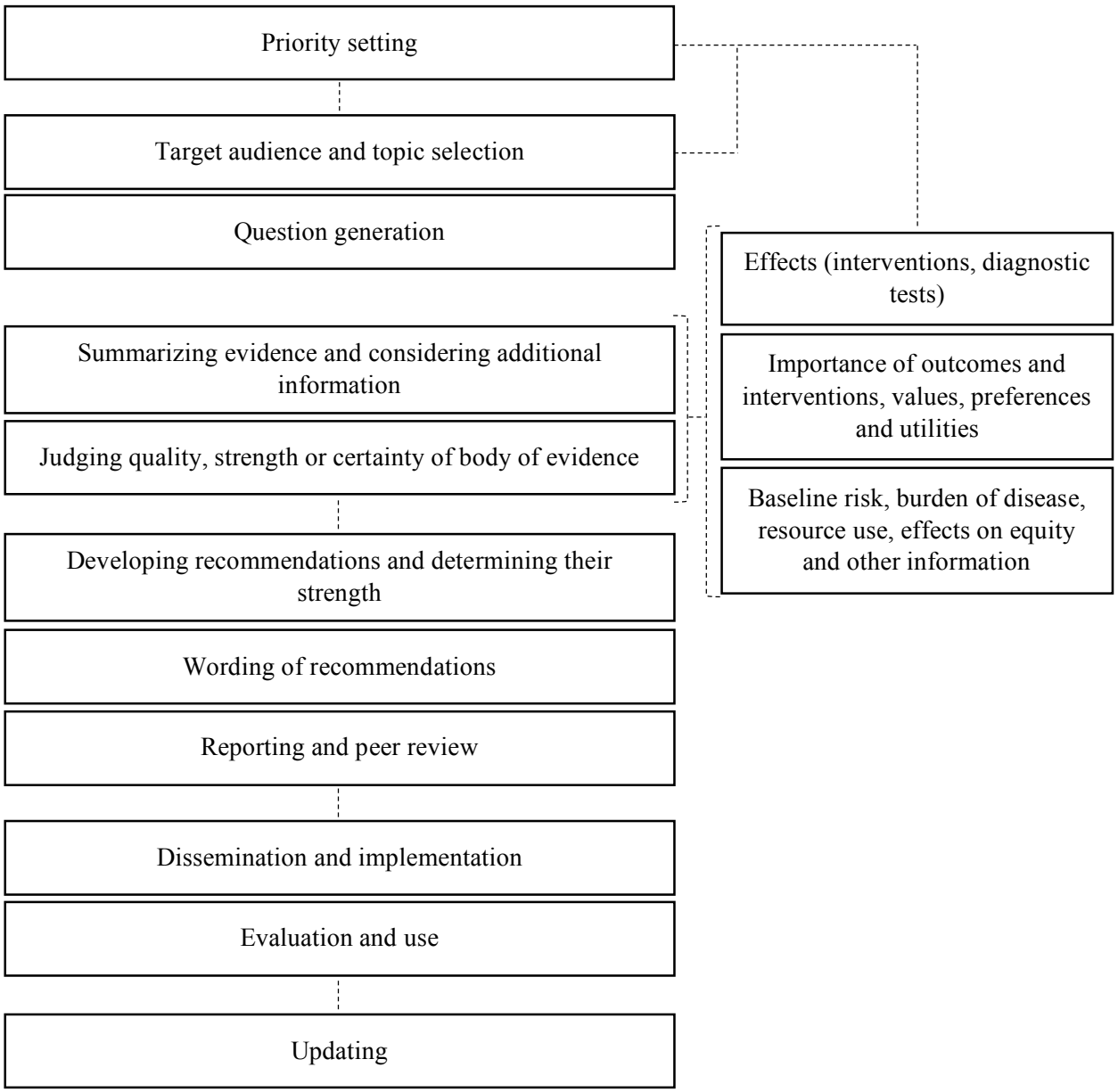


Figure 2. Steps in the guideline development process. Adapted from Schünemann et al. [114].

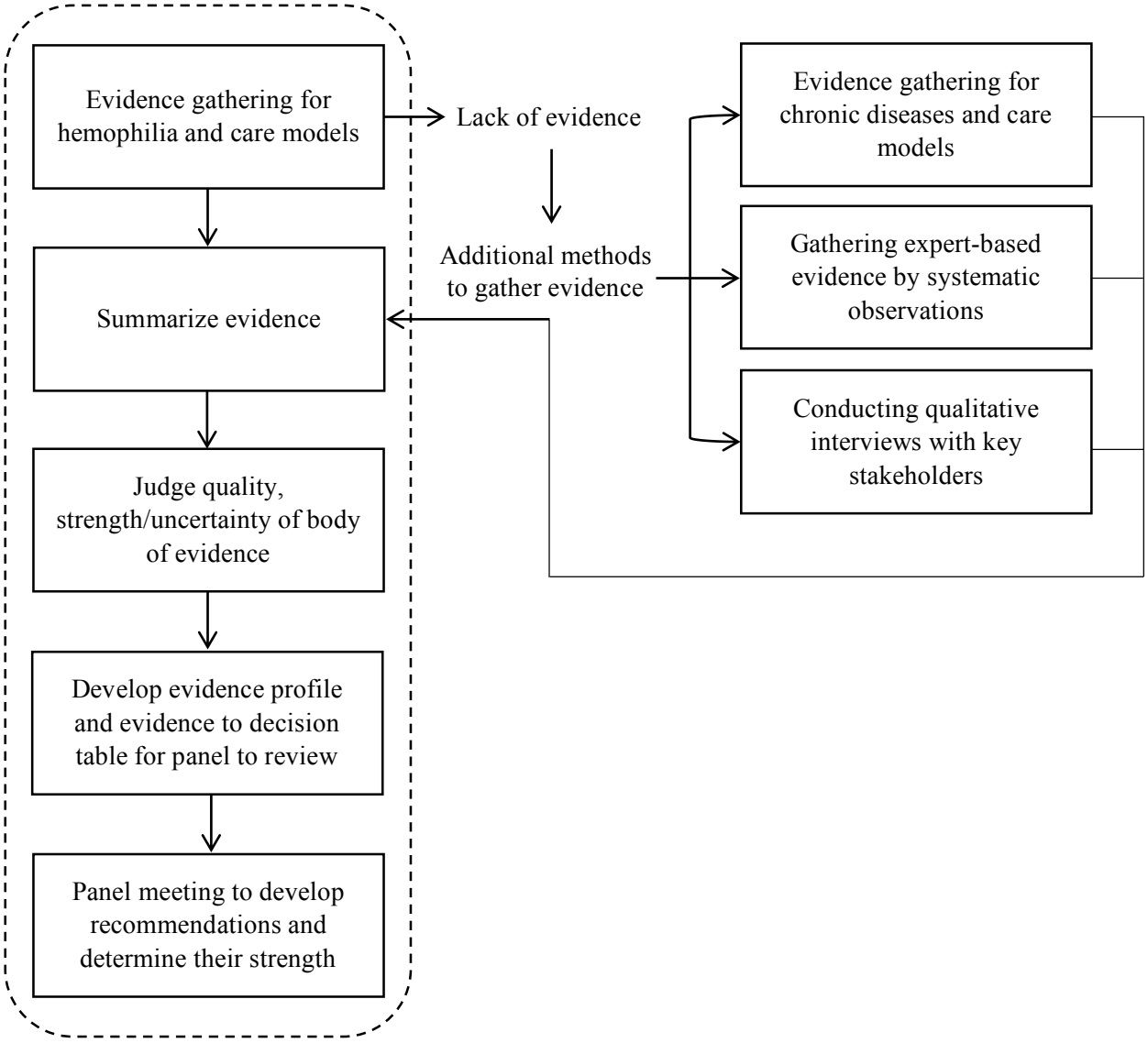


Figure 3. Overview of the methodology to retrieve, evaluate, and synthesize evidence for the NHF-McMaster Guideline on Care Models for Hemophilia Management. The dashed line represent the usual course of guideline development.

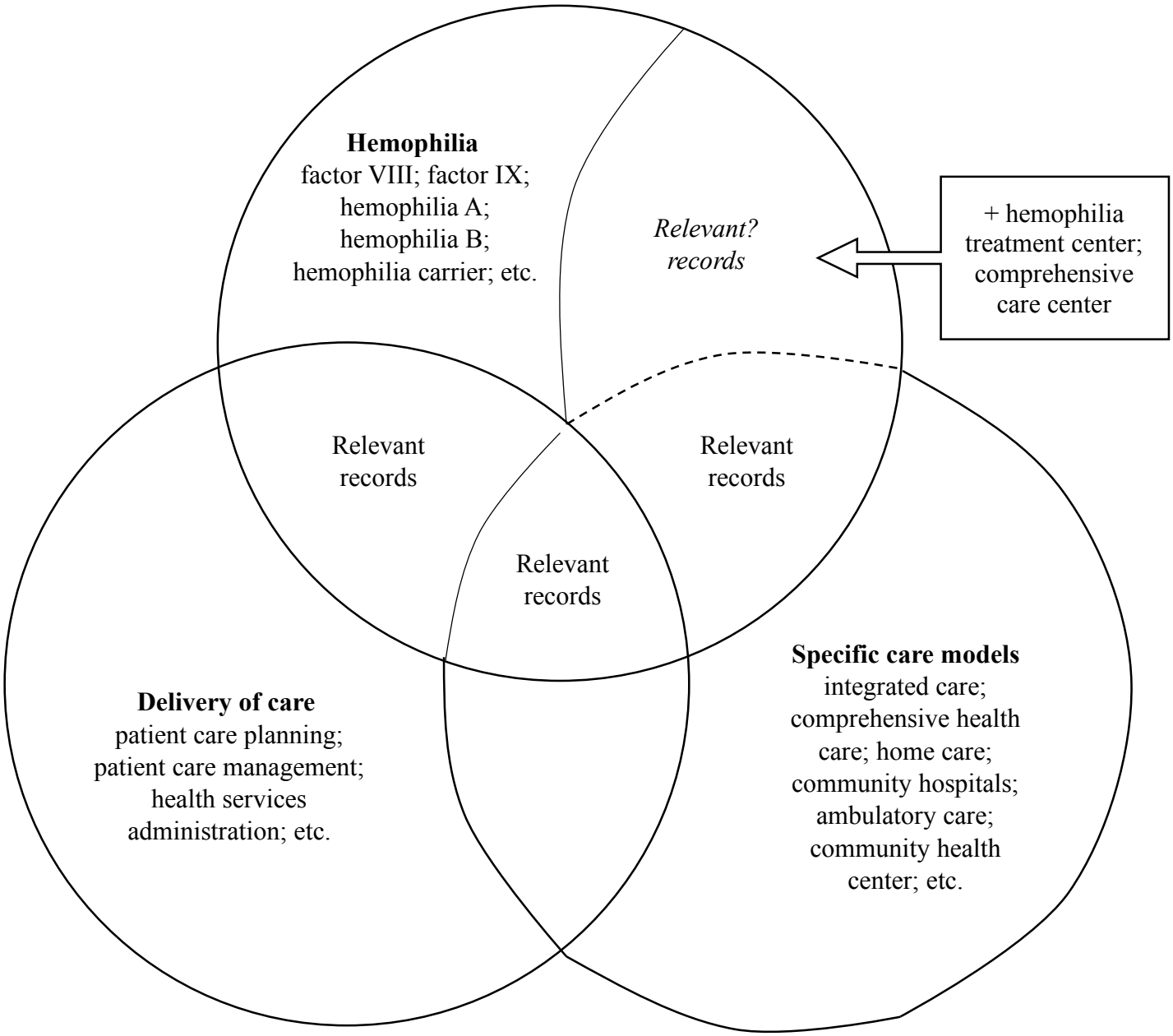


Figure 4. The influence of adding comprehensive care center and hemophilia treatment center terms to the search strategy

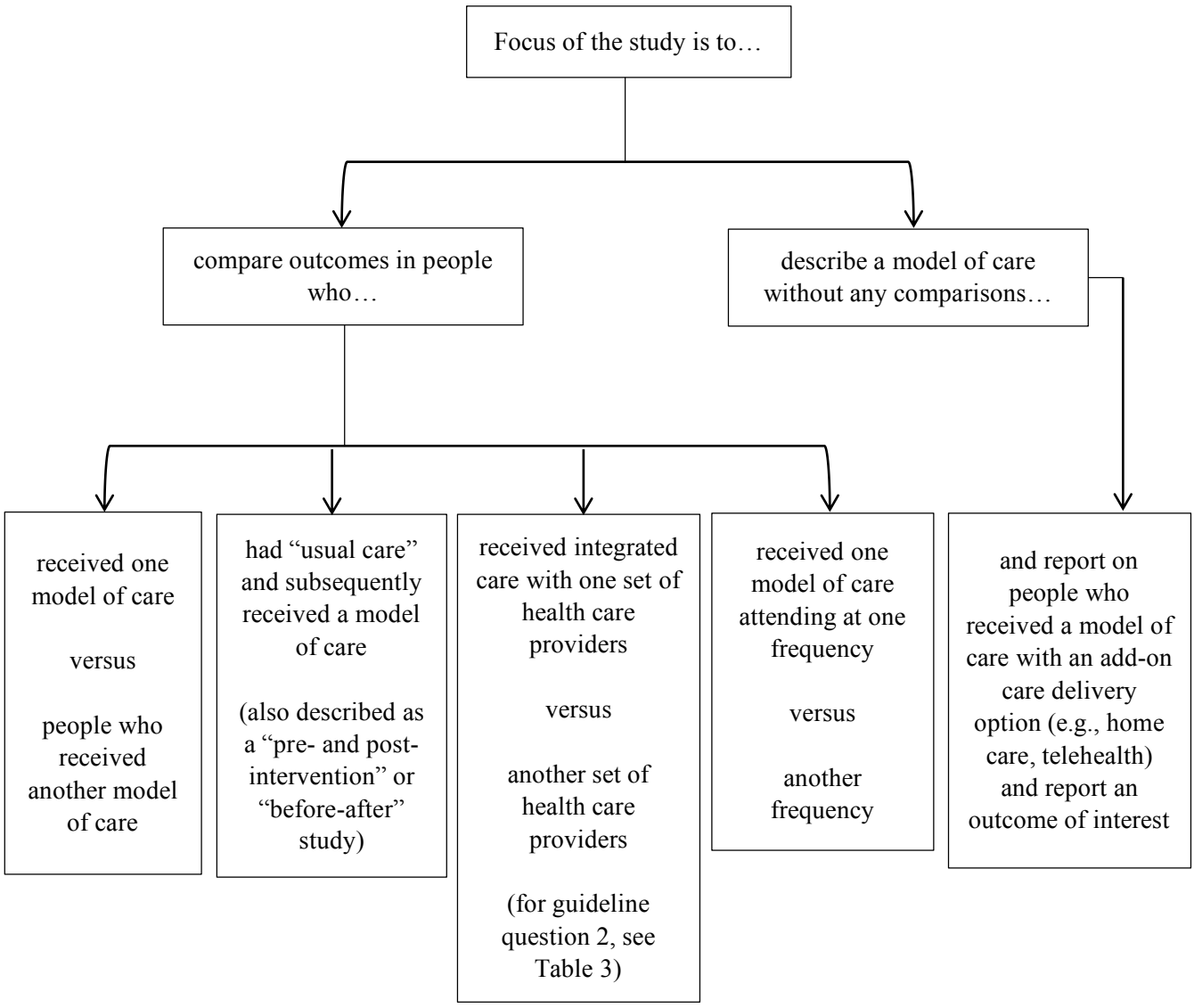


Figure 5. Classification system of hemophilia models of care to reduce screening inclusion criteria ambiguity.

Figure 6. PRISMA flow diagram for systematic review on care models in the management of hemophilia

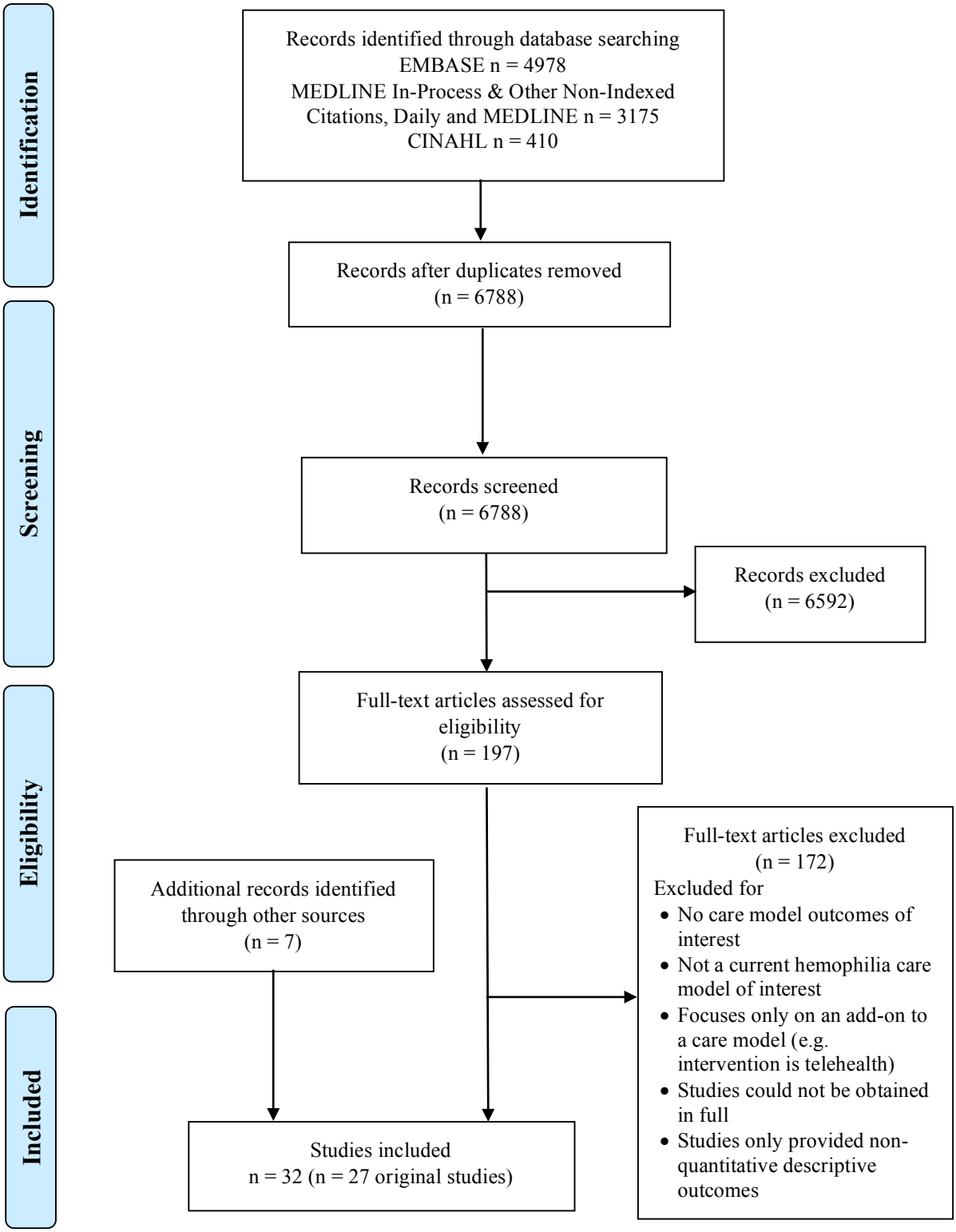


Figure 7. PRISMA flow diagram for overview of reviews on integrated multidisciplinary care for the management of chronic conditions in adults

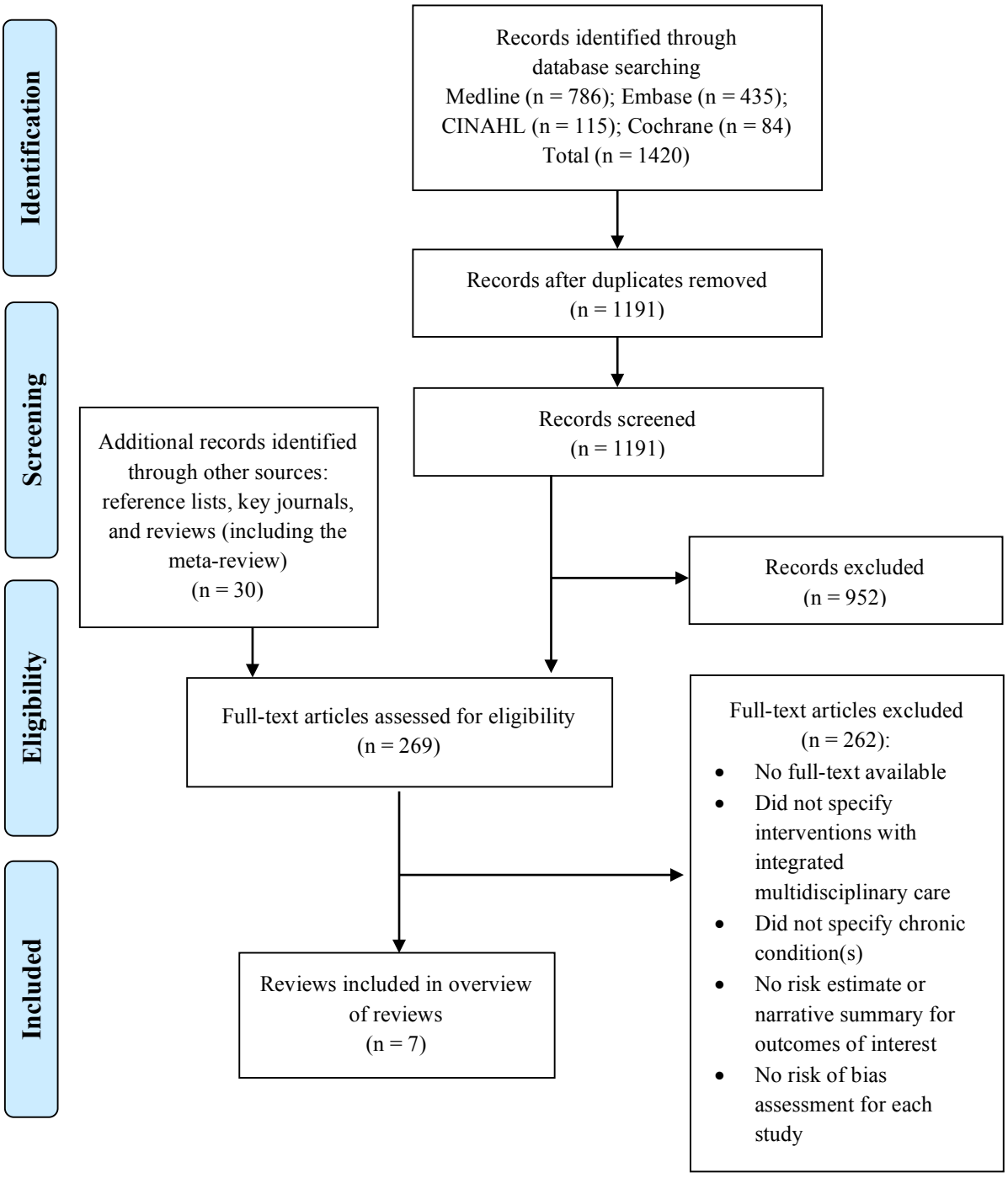


Table 1. Facilitators of and barriers to clinical care, research, and guideline development in rare diseases. This table was developed by Pai *et al*. [115].

Phase	Facilitators	Barriers

Clinical Care	• Though individual rare diseases have low prevalence, total number of patients with rare diseases is large
	• In some jurisdictions, public funds available to ensure access to therapy
	• Existence of European reference networks, centres of expertise, and patient associations
		• Not an actual priority for policymakers, funders
		• Lack of clinical expertise
		• Limited therapeutic options
		  ○ Low availability and/or accessibility
		  ○ Patient eligibility for treatment, dosing forms and administration guidelines vary worldwide
		  ○ Patients and health care providers may be willing to accept treatments with greater risk and unclear benefits

Research	• Though individual rare diseases have a low prevalence, total number of patients with rare diseases is large
	• In some jurisdictions, public funds available for research
	• Existence of European reference networks, centres of expertise, and patient associations
		• General lack of public awareness
		• Not an actual priority for funders, researchers
		• Perceived lack of clinical equipoise
		  ○ Patients and health care providers may be willing to accept treatments with greater risk and unclear benefits
		  ○ Patients and health care providers may be unwilling to accept placebo or comparator treatment
		  ○ Impossible to calculate relative treatment effects if studies are single-arm (i.e., do not have a comparator)
		• Cannot control study results for baseline effects, as these are often unknown
		• Heterogeneity in studies
		• Aggregating data extremely challenging
		• Study enrolment difficult
		  ○ Many rare diseases do not have clear diagnostic criteria
		  ○ Patients not registered in databases in reliable, harmonised way
		  ○ Dearth of epidemiologists and trialists capable of executing creative, methodologically sound studies for rare diseases

Guideline development	• In some jurisdictions, public funds available for methodologic	• Paucity of published data on rare diseases (and much of it is low quality)

	research in guideline creation (e.g. RARE-Bestpractices) • Increasing uptake of GRADE system to summarize evidence, grade its quality, and transparently interpret it to make clinical recommendations • European Union directive on application of patients' rights in cross border healthcare supports European reference networks, which must have capacity to produce good practice guidelines	• Often no published evidence at all for critically important outcomes, or for patient values and preferences

Table 2. NHF-McMaster Guideline on Care Models for Hemophilia Management question 1.

Population	**Intervention**	**Comparator**	**Outcome**
In individuals…	What is the impact of…	Versus…	On…
with severe hemophilia; with non-severe hemophilia; who are carriers of hemophilia Subgroups: comorbidities (inhibitors, infection); access to care; age (pediatric, older population)	integrated care	non-integrated care, including: specialist in a non-specialized care setting; non-specialist in non-specialized care setting; no care	mortality or survival; missed days from work or school; number of emergency room visits; length of in-patient stay; joint damage or disease; quality of life; educational attainment; patient knowledge; patient adherence

Table 3. NHF-McMaster Guideline on Care Models for Hemophilia Management question 2.

Population	**Intervention**	**Comparator**	**Outcome**
In individuals…	What is the impact of…	Versus…	On…
with severe hemophilia; with non-severe hemophilia; who are carriers of hemophilia Subgroups: comorbidities (inhibitors, infection); access to care; age (pediatric, older population)	hematologist or physician with specialisation in hemophilia; nurse with specialisation in hemophilia; physical therapist; social worker; round-the-clock access to a specialised coagulation laboratory	not having each member on the hemophilia care team	mortality or survival; missed days from work or school; number of emergency room visits; length of in-patient stay; joint damage or disease; quality of life; educational attainment; patient knowledge; patient adherence

Table 4. Examples of retrieved studies organized by the developed hemophilia care models classification system and data use for the systematic review.

Focus of the study is to…	Study examples	Description of study	Examples of data used for the systematic review
compare outcomes in people who…			
received one model of care versus people who received another model of care	Soucie *et al*. 2000 [54]	Study of 2,950 patients with hemophilia from six states of the US. Reported outcomes (mortality) in patients receiving care from HTCs compared to those primarily receiving care from non-HTCs (private physicians, hematologists, nonhospital-based clinics, etc.)	Data from patients receiving care from HTCs and data from patients receiving care from non-HTCs were used.
had "usual care" and subsequently received a model of care (also described as a "pre- and post-intervention" or "before and after" study)	Smith and Levine 1984 [52]	Study of hemophilia patients from 11 federally funded Comprehensive Hemophilia Centers. Reported outcomes (missed days of school or work and length of in-patient stay) in 4,742 patients in their fifth year following the initiation of the program, and 2,112 patients in the year preceding the program.	Data from patients in their fifth year following the initiation of the program and data from patients in the year preceding the program were used.
received integrated care with one set of health care providers versus another set of health care providers	We did not find any studies that could be classified in this category.		
received one model of care attending at one frequency versus another frequency	Monahan *et al*. 2011 [45]	Study of 6,420 hemophilia patients from 130 HTCs in the US. Reported outcomes (missed days of school or work, joint damage or disease (or other measures of functional status)) in patients who were frequent users and infrequent users.	Data from patients who were frequent users (one or more visits per year) and data from infrequent users (less than one visit per year, excluding the first visit) were used.
describe a model of care without any comparisons…			

and report on people who received a model of care that may have an add-on care delivery option (e.g., home care, telehealth) and report an outcome of interest	Levine and Britten 1973 [55, 77]	Study of 45 pediatric patients with hemophilia from a single medical center. Studied patients before and after a 12-month home infusion program. Outcomes (mortality, missed days of school or work, and length of in-patient stay) were examined while an integrated care model was in place.	Data from patients after the home-infusion were used. Where an add-on care delivery option (home care) was present, this data were used as it represented the most current state of integrated care.
	Weiss *et al*. 1991 [56]	Study of 166 patients with congenital bleeding disorders (hemophilia, von Willebrand disease) at a single medical center followed for one year. Outcomes (patient adherence) were examined while an integrated care model was in place.	Data from patients retrieved during the one year of follow-up were used. Some patients received home care treatment.

HTC: Hemophilia Treatment Center

Table 5. Description of included non-randomized comparative studies

Study	Design	n	Country, sites	Population	Intervention	Control	Outcome(s) of interest
Arnold *et al*. 2014 [43]	Cross-sectional Surveyed individuals identify their existing knowledge levels and gaps	104	Canada, three HTCs (from Eastern, Central and Western parts of Canada)	All patients were ≥18 years old 49% mild 13% moderate 29% severe (similar to Canadian distribution)	HTC attendance within the past 12 months	No HTC attendance within the past 12 months	Patient knowledge: Knowledge seeking over the past 12 months
Lazerson 1972 [53]	Before-after Patients compared prior to and after development of a comprehensive care center	20 before 20 after	United States, NR (presumably, the Children's Hospital at Stanford)	10 patients were 5-9 years old; 10 patients were 10-17 years old All severe No inhibitors	The year 1970-1971 is the period during which all children were well established in the comprehensive care program	The year 1968 to 1969 was prior to the establishment of a comprehensive care program	Missed days of school or work: Number of days lost from school or work
Monahan *et al*. 2011 [45]	Cross-sectional Universal Data Collection (UDC) data	6420	United States, ~130 HTCs	All patients were ≤18 years old 50.2% severe 24.3% moderate 25.4% mild 14% reported having inhibitors	Patients who were frequent users (one or more visits per year)	Patients who were infrequent users (less than one visit per year, excluding the 1st visit) or only had a 1st visit to the HTC	Missed days of school or work: >11 days lost from school or work per year Joint damage or disease (and other measures of functional status): Decreased activity (work,

							school, recreational, and self care) per year
Smith *et al*. 1982 [51]	Before-after Patients compared prior to and after development of a Comprehensive Hemophilia Center	23 before 43 after	Hemophilia Center of Rhode Island, Rhode Island Hospital, Rhode Island, US	49% ≤16 years old 51% >16 years old 70% severe 30% moderately severe	Three years following the initiation of the Comprehensive Hemophilia Centre program	The year preceding the Comprehensive Hemophilia Centre program	Emergency visits: Number of visits to emergency room and walk-in clinic
Smith and Levine 1984 [52]	Before-after Patients compared prior to and after development of a Comprehensive Hemophilia Center	2,112 before 4,742 after	United States, 11 federally funded Comprehensive Hemophilia Centers	67% severe 33% mild or moderate	The fifth year following the initiation of the Comprehensive Hemophilia Centre program	The year preceding the Comprehensive Hemophilia Centre program	Missed days of school or work: Number of days lost from school or work Length of in-patient stay: Number of days spent as in-patient
Soucie *et al*. 2000 [54]	Prospective cohort Hemophilia Surveillance System (HSS) data	2950	United States, HTCs in Colorado, Georgia, Louisiana, Massachusetts, New York, and Oklahoma	46% 0-19 years old 49% 20-59 years old 5% 60-70+ years old 42% severe 24% moderate 31% mild 5% reported having inhibitors	Patients receiving care in HTCs	Patients received care primarily from private physicians or hematologists, hospital- and nonhospital-based clinics, only from hospitals or emergency rooms, or care from a variety of other sources	Mortality or survival: Mortality adjusted

				2% reporting having liver disease 25% reported having a positive HIV serostatus; 7% AIDS			
Soucie *et al*. 2001 [86]	Prospective cohort Hemophilia Surveillance System (HSS) data	2546	United States, HTCs in Colorado, Georgia, Louisiana, Massachusetts, New York, and Oklahoma	0-24 years old 41% 25-44 years old 45.4% ≥45 years old 13.7% 47% severe 23% moderate 28% mild 5.3% reported having inhibitors	Patients receiving care in HTCs	Patients received care primarily from private physicians or hematologists, hospital- and nonhospital-based clinics, only from hospitals or emergency rooms, or care from a variety of other sources	Emergency room visits: Number of people with at least one hospitalization over four years adjusted
Soucie *et al*. 2004 [71]	Cross-sectional Universal Data Collection (UDC) data	4343	United States, ~130 HTCs	All patients were ≤19 years old 21% mild 24% moderate 55% severe 10.8% reported having inhibitors	Patients who were frequent users (one or more visits per year)	Infrequent users (less than one visit per year)	Joint damage or disease (and other measures of functional status): Overall joint ROM

HTC: Hemophilia Treatment Center; **ROM:** range-of-motion

Table 6. Risk of bias summary by non-randomized comparative study assessed by the ACROBAT-NRSI tool

Study	Bias due to confounding	Bias in selection of participants into the study	Bias in measurement of interventions	Bias due to departures intended interventions	Bias due to missing data	Bias in measurement of outcomes	Bias in selection of reported results
Arnold *et al*. 2014 [43]	Moderate risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
Lazerson 1972 [53]	Serious risk	Low risk	Low risk	Low risk	Serious risk	Low risk	Low risk
Monahan *et al*. 2011 [45]	Moderate risk	Moderate risk	Low risk	Low risk	Low risk	Low risk	Low risk
Smith *et al*. 1982 [51]	Serious risk	Moderate risk	Low risk	Low risk	Unclear risk	Unclear risk	Low risk
Smith and Levine 1984 [52]	Serious risk	Low risk	Low risk	Low risk	Unclear risk	Unclear risk	Low risk
Soucie *et al*. 2000 [54]	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
Soucie *et al*. 2001 [86]	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
Soucie *et al*. 2004 [71]	Moderate risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk

Table 7. GRADE evidence profile for summary of findings from non-randomized comparative studies

Quality assessment							№ of patients		Effect		Quality
№ of studies (participants)	**Study design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Integrated care model**	**Non-integrated care model**	**Relative (95% CI)**	**Absolute (95% CI)**	
Mortality											
1 (2,950)	non-randomised comparative study	not serious [1]	not serious	not serious	not serious	none	149/1979 (7.5%)	86/971 (8.9%)	**RR 0.6** (0.5 to 0.8)	**35 fewer per 1000** (from 18 fewer to 44 fewer)	⊕⊕○○ LOW
Missed days of school or work											
3 (3,032 without and 10,282 with integrated care)	non-randomised comparative study	serious [2]	not serious	not serious [2]	serious [2]	none	4742	2112	-	MD **10.2 lower** (not reported)	⊕○○○ VERY LOW
Emergency room visits											
2 (662 without and 1,950 with integrated care)	non-randomised comparative study	not serious [1]	not serious	not serious [3]	not serious	none	557/1907 (29.2%)	225/639 (35.2%)	**RR 0.6** (0.5 to 0.7)	**141 fewer per 1000** (from 106 fewer to 176 fewer)	⊕⊕○○ LOW
Length of in-patient stay											

Quality assessment							№ of patients		Effect		Quality
№ of studies (participants)	**Study design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Integrated care model**	**Non-integrated care model**	**Relative (95% CI)**	**Absolute (95% CI)**	
1 (4,742 without and with integrated care)	non-randomized comparative study	serious [2]	not serious	not serious [2]	serious [2]	none	4742	2112	-	MD **7.6 fewer** (not reported)	⊕○○○ VERY LOW
Quality of life – not measured											
Joint damage or disease (and other measures of functional status)											
2 (10,763)	non-randomized comparative study	not serious [4]	not serious	serious [4]	not serious	none	Severe disease: frequent HTC users had less ROM limitation than infrequent users (unadjusted analysis). Moderate disease: infrequent HTC users has less ROM limitation than frequent users (adjusted analysis). Mild disease: infrequent HTC users has less ROM limitation than frequent users (adjusted analysis).				⊕○○○ VERY LOW
Educational attainment – not measured											
Patient knowledge											
1 (104)	non-randomized comparative study	not serious [5]	not serious	serious [5]	serious [5]	none	Significantly fewer people who did not attend HTC in past 12 months sought information				⊕○○○ VERY LOW
Patient adherence – not measured											

CI: confidence interval; **RR:** risk ratio; **MD:** mean difference; **HTC**: Hemophilia Treatment Center; **ROM**: range-of-motion

1. Potential for bias related to definition of integrated care as an HTC user with at least one visit to the center but not downgraded.

2. Overall, the results were not adjusted for confounding factors, and the 95% confidence intervals were not calculated. There is also some indirectness as the integrated care model in 1970s/1980s has changed.
3. Not downgraded, although the number of hospitalizations was used as a surrogate for number of emergency room visits
4. Overall, downgraded once for unadjusted analysis depending on severity of disease and differences in definition of integrated care and non-integrated care by frequency of HTC use
5. Overall, the results were not adjusted for confounding, integrated care was defined as attendance at an HTC in last 12 months, and there were few events.

Table 8. Outcome data from non-randomized non-comparative studies

Mortality outcome data from non-comparative studies

Study	Number of events (deaths)	Sample size	Event rate (mortality)	Standard error
Levine and Britten 1973 [55, 77]	1	45	0.02	0.02
Strawczynski *et al*. 1973 [71]	1	40	0.03	0.03
Isarangkura *et al*. 1987 [57]	1	10	0.10	0.10
Kennelly *et al*. 1995 [75]	7	60	0.12	0.04
Chuansumrit *et al*. 1999 [74]	6	96	0.06	0.03
Mahlangu *et al*. 2009 [73]	9	1451	0.01	0.002
Range of mortality event rate by the number of deaths: 6 to 100 deaths per 1000 persons				

Missed days of school or work outcome data from non-comparative studies

Study	Sample size	Mean (days/person/year)[a]
Rabiner *et al*. 1972 [79, 80]	13	25.85
Levine and Britten 1973 [53, 72]	41	6.80
Strawczynski *et al*. 1973 [76]	36	15.00
Hilgartner 1977 [81]	17	0.70
Panicucci *et al*. 1977 [49, 50]	64	15.20
Ekert *et al*. 1981 [79, 80]	24	10.00[b]
Szucs *et al*. 1998 [82, 83]	566	54.76
Range of mean number of days missed: 0.7 to 55 days per patient per year		

[a]standard deviation could not be calculated and/or was not reported in the study; [b]reported as a median and not included in the mean number of days per patient per year range

Number of emergency room visits outcome data from non-comparative studies

Study	Sample size	Mean (visits/person/year)[a]
Panicucci *et al*. 1977 [49, 50]	64	0.09
Tencer *et al*. 2007 [87]	34	0.50
Range of mean number of emergency room visits: 9 to 500 per 1000 persons per year		

[a]standard deviation could not be calculated and/or was not reported in the study

Length of in-patient stay outcome data from non-comparative studies*

Study	Sample size	Mean (days/person/year)[a]
Levine and Britten 1973 [55, 77]	45	0.95
Strawczynski *et al*. 1973 [53, 72]	36	6.69
McKenzie *et al*. 1974 [92]	5	0.80[b]
Carter *et al*. 1976 [89]	114	5.05
Kennelly *et al*. 1995 [75]	60	14.50
Szucs *et al*. 1998 [82, 83]	566	1.00

Chuansumrit *et al*. 1999 [74]	96	7.90
Martínez-Murillo *et al*. 2004 [116]	74	1.90
Heemstra *et al*. 2005 [90]	17	2.01
Tencer *et al*. 2007 [87]	34	0.40
Range of mean number of days of patient stay: 0.4 to 14.5 days per person per year		

[a]standard deviation could not be calculated and/or was not reported in the study; [b]standard deviation was available, 1.1 days per patient per year

Joint damage or disease (or other measures of functional status) outcome data from non-comparative studies

Study	Number of events or persons[a]	Total number of persons	Event rate[b] or proportion of persons with joint damage or disease	Standard error
Joint damage or disease (per joint)				
Ekert *et al*. 1981 [79, 80]	45	192	0.23	0.03
Isarangkura *et al*. 1987 [57]	3	9	0.33	0.16
Range of joint damage or disease event rate by the number of events: 234 to 333 per 1000 persons				
Joint damage or disease (per person)				
Rabiner *et al*. 1972 [84, 85]	1	16	0.06	0.06
Levine and Britten 1973 [55, 77]	2	45	0.04	0.03
Hilgartner *et al*. 1977 [81]	8	68	0.12	0.04
Ingram *et al*. 1979 [93]	12	28	0.43	0.09
Range of proportion of persons with joint damage or disease: 44 to 429 persons per 1000 persons				

[a]joints damaged or diseased; [b]joints damaged or diseased per person

Patient adherence outcome data from non-comparative studies

Study	Number of adhering persons	Total number of persons	Proportion of adhering persons	Standard error
Hilgartner *et al*. 1977 [81]	6	72	0.08	0.03
Weiss *et al*. 1991 [56]	82	166	0.49	0.04
Range of proportion of adhering persons: 83 to 494 per 1000 persons				

Table 9. Characteristics of the included reviews and studies of the overview of reviews

Review, date assessed up to date	**Population and intervention of the review**	**Population of the included studies of the review**	**Intervention of the included studies of the review**	**Outcomes of the included studies of the review (follow-up)**
Baishnab and Karner 2012 01/12/2011	Patients with **asthma** of any age. Primary care based practices offering a proactive system of care by organised **asthma clinics**, and practices that undertook shared care with hospital services.	Children (from one out of the three studies) and adults (mean age 27 to 48 years) with **asthma**.	Studies published from 1999 to 2004. Care provided in an asthma clinic commonly included assessment, education, counselling, and management, which were delivered primarily by nurses and general practitioners.	Missed days of school or work (6 months) Emergency room visits (range: 6-9 months) Quality of life (SGRQ, AQLQ, EQ4D) (range: 4-9 months)
Health Quality Ontario 2009 03/10/2008 (search date)	Patients with **heart failure**. **Disease management programs** with multidisciplinary approaches with a team consisting of a nurse and physician, one of which is a specialist in heart failure management. Program begins after discharge from the hospital.	Adults with **heart failure**, mean age 65 to 78 years.	Studies published from 2002 to 2007. Care commonly included services to manage disease through formalized links between primary and specialized care (e.g. direct or telephone) with patient education, diet and activity counselling, self-management primarily by a heart failure physician or cardiologist, nurse, dietitian, and involvement of the general practitioner. Other care providers could include a psychologist and physiotherapist.	Mortality (12 months) Emergency room visits (12 months) Length of in-patient stay (mean: 12 months) Quality of life (MLHFQ) (mean: 12 months)
Kruis *et al*. 2013 12/04/2012	Patients with **COPD**.	Adults with **COPD**, mean age 62 to 77 years.	Studies published from 1994 to 2011. Care programs that commonly include patient	Mortality (mean: 12 months)

	Integrated disease management including an active involvement of at least two healthcare providers; and organizational, professional, patient-directed and financial interventions.		education, self-management, monitoring, physical training, and nutrition advice by care teams that could include a nurse, physiotherapist, dietician, psychologist, occupational therapist, social worker, and pulmonologist.	Emergency room visits (mean: 12 months) Length of in-patient stay (mean: 12 months) Quality of life (SGRQ) (range: 3-24 months) Functional status (6MWD) (range: 3-12 months)
Lemmens *et al*. 2009 05/2008	Patients ≥16 years with **asthma** or **COPD**. **Integrated disease management** including patient-related, professional-directed, and organisational interventions.	Adults with **COPD** (or **asthma**), mean age 48 to 80 years.	Studies published from 2002 to 2006. Care commonly included patient and provider education, and continuity of care (hospital to home with case management or follow-up at home), primarily by nurse and/or general practitioner.	Emergency room visits (range: 2-9 months) Quality of life (SGRQ) (1 of 6 studies from asthma) (range: 3-12 months)
		Adults with **asthma**, mean age 35 to 45 years.	Studies published from 1998 to 2008. Care commonly included pharmacy programmes with patient education, monitoring, and expansion or revision of professional roles (pharmacist plays more of active role).	Quality of life (AQLQ) (range: 3-12 months)
Peytremann-Bridevaux *et al*. 2008	Adults with **COPD**.	Adults with **COPD**, mean age 62 to 74 years.	Studies published from 1995 to 2006. Care commonly included patient education,	Mortality (range: 12-24 months)

12/2006	**Disease management** including two or more different components (e.g., physical exercise, self-management, structured follow-up), two or more health care professionals actively involved in patient care, patient education, and at least one component of the intervention lasted a minimum of 12 months.		self-management, exercise, structured follow-up, involving two or more health care professionals (e.g., general practitioner, pulmonary care physician, physiotherapist).	Functional status (6MWD) (range: 12-24 months)
Peytremann-Bridevaux *et al*. 2015 06/2014	Adults ≥16 years with **asthma**. **Disease management** with at least one organizational component targeting patients and at least one targeting healthcare professionals or the healthcare system; patient education or self-management support component; active involvement of two or more healthcare professionals in patient care; and minimum duration of three months for at least one component.	Adults with **asthma**, mean age 36 to 49 years.	Studies published from 2000 to 2010. Care programmes that commonly included structured follow-up, teamwork between heath care providers (general practitioner, nurse, pharmacist, care manager), patient and provider education, self-management support.	Missed days of school or work (12 months) Emergency room visits (range: mean 6 months) Quality of life (AQLQ) (range: 6-12 months)
Takeda *et al*. 2012 01/2009	Patients ≥18 years with **heart failure**. **Clinical service interventions** (inpatient,	People with **heart failure**, mean age 72 to 78 years.	Studies published from 1999 to 2008. Care commonly included interventions of home care usually with telephone follow-up, an	Mortality (range: 6 to 12 months) Quality of life (not reported by review)

	outpatient or community-based interventions or packages of care). These interventions included: case management, clinical interventions, and multidisciplinary interventions.	information program with education, a treatment plan with diet and exercise management, and increased communication between heart failure nurses, cardiologist, primary care physicians, and other health care providers.

SGRQ: St. George's Respiratory Questionnaire; **AQLQ:** Asthma Quality of Life Questionnaire; **EQ4D:** EuroQol four dimensions questionnaire; **MLHFQ:** Minnesota Living with Heart Failure Questionnaire; **COPD:** chronic obstructive pulmonary disease; **IDM:** integrated disease management; **6MWT:** Six Minute Walking Test

Table 10. GRADE summary of findings table of all reviews by outcome and by chronic condition

Chronic condition, Review	**Anticipated absolute effects** (95% CI)		**Relative effect** (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)
	Risk with non-integrated multidisciplinary care	**Risk with integrated multidisciplinary care**			
Mortality					
COPD, Kruis *et al*. 2013 follow up: mean 12 months	185 per 1000	**162 per 1000** (100 to 249)	**OR 0.85** (0.49 to 1.46)	1235 (5 RCTs)	⊕⊕⊕◯ MODERATE [1,2,3]
COPD, Peytremann-Bridevaux *et al*. 2008 follow up: mean 12 months	89 per 1000	**77 per 1000** (50 to 118)	**OR 0.85** (0.54 to 1.36)	1102 (9 RCTs, 1 nRCT)	⊕⊕⊕◯ MODERATE [3,4]
Heart failure, Health Quality Ontario 2009 follow up: mean 12 months	194 per 1000	**138 per 1000** (109 to 177)	**RR 0.71** (0.56 to 0.91)	2787 (8 RCTs)	⊕⊕⊕◯ MODERATE [2,4,5]
Heart failure, Takeda *et al*. 2012 follow up: range 6 to 12 months	195 per 1000	**127 per 1000** (95 to 169)	**OR 0.60** (0.43 to 0.84)	1784 (8 RCTs)	⊕⊕⊕⊕ HIGH [2]
Missed days of school or work					
Asthma, Baishnab and Karner 2012 assessed with: days lost per person follow up: 6 months	The mean missed days of school or work was **5.37** days	The mean missed days of school or work was 0.57 days fewer (1.57 fewer to 0.48 more)	-	191 (1 RCT)	⊕⊕◯◯ LOW [3,7]
Asthma, Peytremann-Bridevaux *et al*. 2015 assessed with: days lost per person follow up: mean 12 months	Two studies reported comparable missed days between the intervention and control.			477 (2 RCTs)	⊕⊕◯◯ LOW [3,8]
Emergency room visits					

COPD, Kruis *et al*. 2013 assessed with: patients with at least one visit follow up: range 3 to 12 months	281 per 1000	**200 per 1000** (114 to 328)	**OR 0.64** (0.33 to 1.25)	1161 (4 RCTs)	⊕⊕⊕◯ MODERATE [1,9]
COPD, Lemmens *et al*. 2009 assessed with: visits per person follow up: range 2 to 9 months	The mean emergency room visits was **0.1 to 1.59** visits	The mean emergency room visits was 0.08 visits fewer (0.03 fewer to 0.18 fewer)	-	462 (4 RCTs)	⊕⊕◯◯ LOW [5,7]
Asthma, Baishnab and Karner 2012 assessed with: patients with at least one visit follow up: range 6 to 9 months	17 per 1000	**18 per 1000** (4 to 82)	**OR 1.03** (0.21 to 5.15)	344 (2 RCTs)	⊕⊕◯◯ LOW [2,3,10]
Asthma, Peytremann-Bridevaux *et al*. 2015 assessed with: visits per person follow up: mean 6 months	One RCT found a reduction in emergency department visits. Four RCTs and one nRCT did not find an effect. One CBA study found no important reduction between groups.			(5 RCTs; 1 nRCT; 1 CBA)	⊕⊕⊕◯ MODERATE [11,12]
Heart failure, Health Quality Ontario 2009 assessed with: visits or hospital readmissions per person follow up: 12 months	840 per 1000	**707 per 1000** (577 to 828)	**HR 0.67** (0.47 to 0.96)	151 (1 RCT)	⊕⊕◯◯ LOW [3,5,13]
Length of in-patient stay					
COPD, Kruis *et al*. 2013 assessed with: days per patient follow up: mean 12 months	The mean length of stay was **1.6 to 18.2** days	The mean length of stay was 3.78 days fewer (1.67 fewer to 5.9 fewer)	-	741 (6 RCTs)	⊕⊕⊕⊕ HIGH
Heart failure, Health Quality Ontario 2009 assessed with: days in hospital or total hospital days per patient follow up: mean 12 months	Overall, appears that persons receiving integrated care had shorter hospital stay, however, only three studies were statistically significant for duration of stay and other study showed longer hospital stays.			(7 RCTs)	⊕⊕◯◯ LOW [4,5,14]
Quality of life					

COPD, Kruis *et al.* 2013 assessed with: change in SGRQ total score Scale from: 0 to 100 (worse) follow up: range 3 to 12 months	The mean change in SGRQ total score was **6.24 to –3.4** points	The mean change in SQRQ total score was 3.71 points lower (5.83 lower to 1.59 lower)	-	1425 (13 RCTs)	⊕⊕⊕⊕ HIGH [2]
COPD (5 of 6 studies) and asthma (1 of 6 studies), Lemmens *et al.* 2009 assessed with: SGRQ total score Scale from: 0 to 100 (worse) follow up: range 3 to 12 months	The mean SGRQ total score was **58.5 to 27.3** points	The mean SGRQ total score was 2.52 points lower (5 to 0.05 lower)	-	770 (6 RCTs)	⊕⊕⊕⊕ HIGH [15]
Asthma, Baishnab and Karner 2012 assessed with: SGRQ, AQLQ, EQ4D total scores follow up: range 4 to 9 months	One study reported a non-significant reduction. The other study reported 0 change between groups.			(2 RCTs)	⊕⊕⊕◯ MODERATE [3,10]
Asthma, Lemmens *et al.* 2009 assessed with: AQLQ total score Scale from: 1 (worse) to 7 follow up: range 3 to 12 months	The mean AQLQ total score was **4.4 to 5.8** points	The mean AQLQ total score was 0.35 points higher (0.21 higher to 0.5 higher)	-	841 (3 RCTs, 1 CBA)	⊕⊕⊕◯ MODERATE [9,11]
Asthma, Peytremann-Bridevaux *et al.* 2015 assessed with: AQLQ and Mini AQLQ Scale from: 1 (worse) to 7 follow up: range 3 to 12 months	-	SMD **0.22 SD higher** (0.08 higher to 0.37 higher)	-	1627 (8 RCTs)	⊕⊕⊕◯ MODERATE [8]
Heart failure, Health Quality Ontario 2009 assessed with: MLHFQ follow up: mean 12 months	Intervention may improve quality of life. Quality of life improvement may be seen in physical domain, but not emotional domain.			(5 RCTs)	⊕⊕◯◯ LOW [16,17]
Heart failure, Takeda *et al.* 2012 assessed with: not reported by review follow up: not reported by review	Heterogeneity in outcome measures and methods of reporting findings. Five of the studies reported very little difference in health-related quality of life scores between groups.			(9 RCTs)	⊕⊕◯◯ LOW [8,18]
Functional status					
COPD, Kruis *et al.* 2013 assessed with: change in the 6MWT follow up: range 3 to 12 months	The mean change in distance was **-38.03 to 46** meters	The mean distance was 43.86 meters more (21.83 more to 65.89 more)	-	871 (14 RCTs)	⊕⊕⊕◯ MODERATE [19]

COPD, Peytremann-Bridevaux *et al.* 2008 assessed with: 6MWT follow up: mean 12 months	The mean distance was **290.3 to 430.8** meters	The mean distance was 32.2 meters more (4.1 more to 60.3 more)	-	359 (5 RCTs)	⊕⊕⊕◯ MODERATE [3,4]
Educational attainment – not measured					
Patient adherence – not measured					
Patient knowledge – not measured					

CI: confidence interval; **OR:** odds ratio; **RR:** risk ratio; **HR**: hazard ratio; **COPD:** chronic obstructive pulmonary disease; **RCT:** randomized controlled trial; **nRCT:** non-randomized controlled trial; **CBA:** controlled before-after study; **SGRQ:** St. George's Respiratory Questionnaire; **AQLQ:** Asthma Quality of Life Questionnaire; **EQ4D:** EuroQol four dimensions questionnaire; **MLHFQ:** Minnesota Living with Heart Failure Questionnaire; **6MWT:** Six Minute Walking Test

1. Not downgraded, but some studies with incomplete outcome reporting or selective reporting.
2. Some inconsistency across studies (low to moderate, i.e., I^2 <40%, or 30-60%).
3. Few events and/or effect includes appreciable harm, no effect, and benefit.
4. Not downgraded, but some studies with no or unclear allocation concealment.
5. Some studies did not involve an active multidisciplinary team in the intervention.
6. Additional study with 72 adults with asthma reported fewer people missing days of work when attending clinics.
7. Unclear concealment, no blinding, unbalanced baseline characteristics due to inadequate randomization.
8. Incomplete outcome data and unclear blinding, selective reporting, and inadequate randomization.
9. Inconsistency across studies (substantial to considerable, i.e., I^2 = 50-90%).
10. Unclear allocation concealment and no blinding.
11. Not downgraded, but some studies with incomplete outcome reporting and unclear or ineffective allocation concealment.
12. Mean number of emergency room visits within intervention and within control groups differed greatly between studies.
13. Outcome as composite of emergency department visits and hospital readmissions, considered with risk of bias (unclear allocation concealment).
14. Wide variation in mean number of days reported by studies and unreported standard deviations.
15. Effect not clinically meaningful (2.5 points on a scale of 0-100), results showed more improvement with more types of interventions provided.
16. Inconsistency between total mean scores and domain scores.
17. Wide confidence intervals for domain scores.
18. Small proportions of patients completing questionnaires at all time points.
19. Inconsistency across studies (I^2 = 83%), explained by quality of studies. Considered with risk of bias from no or unclear allocation concealment. Number of meters improved was not clinically relevant, only 15.5 meters (6.37 to 23.93), when restricted to high quality studies.

Table 11. GRADE evidence profile for care models in hemophilia (data from other chronic conditions)

Quality assessment							№ of patients		Effect		Quality
Condition, Review	**Study design (№ of studies)**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Integrated multi-disciplinary care**	**Non-integrated multi-disciplinary care**	**Relative (95% CI)**	**Absolute (95% CI)**	
Mortality (follow up: range 3 to 24 months)											
COPD Kruis *et al*. 2013	randomized trials (5)	not serious [1]	not serious [2]	not serious [3]	serious [4]	none	101/614 (16.4%)	115/621 (18.5%)	**OR 0.85** (0.49 to 1.46)	**23 fewer per 1000** (from 64 more to 85 fewer)	⊕⊕⊕◯ MODERATE
Missed days of school or work (follow up: 6 months; measured: days lost per person)											
Asthma Baishnab and Karner 2012	randomized trials (1)	serious [5]	not serious	serious [3]	serious [4]	none	97	94	-	MD **0.57 fewer** (1.57 fewer to 0.48 more) [6]	⊕◯◯◯ VERY LOW
Emergency room visits* (follow up: range 2 to 9 months; measured: number of visits per person)											
COPD Lemmens *et al*. 2009	randomized trials (4)	serious [5]	not serious	not serious [3]	not serious	none	248	214	-	MD **0.08 fewer** (0.18 fewer to 0.03 fewer)	⊕⊕⊕◯ MODERATE
Length of in-patient stay* (follow up: range 3 to 12 months; measured: days per person)											

Quality assessment							№ of patients		Effect		Quality
Condition, Review	**Study design (№ of studies)**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Integrated multi-disciplinary care**	**Non-integrated multi-disciplinary care**	**Relative (95% CI)**	**Absolute (95% CI)**	
COPD Kruis *et al*. 2013	randomized trials (6)	not serious [7]	not serious	serious [3]	not serious	none	382	359	-	MD **3.78 fewer** (5.9 fewer to 1.67 fewer)	⊕⊕⊕◯ MODERATE
Quality of life* (follow up: range 3 to 12 months; assessed with: SGRQ total score (range 0 to 100 (worse)))											
Asthma (1 of 6 studies), COPD (5 of 6 studies) Lemmens *et al*. 2009	randomized trials (6)	not serious [7]	not serious	serious [3]	not serious	none	718	707	-	MD **2.52 lower** [8] (0 to 5 lower)	⊕⊕⊕◯ MODERATE
Functional status measured as exercise capacity* (follow up: range 3 to 12 months; assessed with: 6 minute walking test (difference is distance in meters))											
COPD Kruis *et al*. 2013	randomized trials (14)	serious [9]	not serious	not serious [3]	not serious	none	466	405	-	MD **43.86 more** (21.83 more to 65.89 more)	⊕⊕⊕◯ MODERATE
Educational attainment - not measured											
Patient adherence - not measured											
Patient knowledge - not measured											

CI: confidence interval; **OR:** odds ratio; **MD:** mean difference; **COPD:** chronic obstructive pulmonary disease; **SGRQ:** St. George's Respiratory Questionnaire; *****Evidence contributed to the NHF-McMaster Guideline on Care Models for Hemophilia Management

1. Some studies with incomplete outcome reporting or selective reporting.
2. Some inconsistency across studies ($I^2 = 55\%$).

3. Studies in people with asthma or COPD were considered to be sufficiently direct. However, some studies downgraded for indirectness based on intervention or outcome.
4. Few events or participants in the studies, or confidence intervals either include appreciable harm, no effect, and benefit.
5. Risk of bias due to unclear concealment, no blinding, and unbalanced baseline characteristics due to inadequate randomization. Some studies risk of bias due to selective reporting.
6. Additional study with 72 adults with asthma showed fewer people missed days of work when attending the clinics.
7. Not downgraded. Analysis of studies with low risk of bias showed similar results.
8. May have little relevance of 2.5 points change on scale of 0-100.
9. Considered with inconsistency across studies (I^2 = 83%), which was explained by the risk of bias - the number of meters improved was 15.5 metres (6.37 to 23.93) when restricted to high quality studies.

Appendix A. Care models for hemophilia management search strategies in MEDLINE, EMBASE and CINAHL

Care models for hemophilia management – MEDLINE

1. exp factor VIII/ or (factor adj VIII).mp.
2. exp factor IX/ or (factor adj IX).mp.
3. hemophilia A/ or hemophilia B/ or (h?emophilia or Christmas disease).mp.
4. (h?emophilia adj2 carrier).mp.
5. 1 or 2 or 3 or 4
6. exp delivery of health care/ or exp patient care planning/ or exp patient care management/ or exp health services administration/ or exp health services accessibility/
7. (delivery of health care or patient care plan* or patient care management or health services administration or health services administration).mp.
8. exp delivery of health care, integrated/ or exp patient care team/ or exp comprehensive health care/
9. (integrated delivery of health care or patient care team or comprehensive health care or comprehensive care cent*).mp. or treatment cent*.tw.
10. ((standard* or contin* or effectiv* or evidence-based or multidisciplin* or integrated or interdisciplin* or collaborat* or model* or ambulatory) adj (care* or network or delivery)).mp.
11. exp home care services/ or home infusion therapy/ or home nursing/
12. hospitals, community/ or ambulatory care/ or community health center/
13. 6 or 7 or 8 or 9 or 10 or 11 or 12
14. 5 and 13

Care models for hemophilia management – EMBASE

1. exp blood clotting factor 8/ or (factor adj VIII).mp.
2. exp blood clotting factor 9/ or (factor adj IX).mp.
3. exp hemophilia/ or exp hemophilia a/ or exp hemophilia b/ or (h?emophilia or christmas disease).mp.
4. 1 or 2 or 3
5. exp health care delivery/ or integrated health care system/ or patient care planning/ or patient care/
6. (health care delivery or integrated health care system or patient care planning or patient care or comprehensive care cent*).mp.
7. ((standard* or contin* or effectiv* or evidence-based or multidisciplin* or interdisciplin* or integrated or collaborat* or model* or ambulatory) adj (care or network or delivery)).mp.
8. exp home care/ or (home infusion therapy or home nursing).mp.
9. community hospital/ or community health cent*.mp.
10. treatment cent*.tw.

11. 4 or 5 or 6 or 7 or 8 or 9 or 10
12. 4 and 11

Care models for hemophilia management – CINAHL

#	Query
S10	S1 AND S9
S9	S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8
S8	TX treatment cent*
S7	(MH "Hospitals") OR (MH "Allied Health Personnel+") OR (MH "Expert Clinicians") OR (MH "Physicians, Family") OR (MH "Home Nursing") OR (MH "Ambulatory Care Nursing") OR (MM "Community Health Nursing") OR (MH "Home Nursing, Professional")
S6	(MH "Home Health Care+") OR MH "Ambulatory Care Facilities+")
S5	(MH "Multidisciplinary Care Team+")
S4	(MH "Health Care Delivery+") OR "health care delivery" OR (MH "Health Care Delivery, Integrated") OR (MH "Health Services Needs and Demand") OR (MH "Health Services Accessibility")
S3	TX network
S2	MW care OR TX care
S1	MH Hemophilia OR TX h?emophilia OR AB h?emophilia OR TI h?emophilia

Appendix B. Description of single arm non-comparative non-randomized studies

Study	Description
Abdullaev *et al*. 1983 [88]	A study of 20 patients (60% were pediatric) with hemophilia at a custodial specialized school, followed for 2-5 years. Outcomes were examined while an integrated care model was in place. Integrated care personnel included a hematologist, nurse, physical therapist, orthopaedic surgeon and a dentist. Integrated care was provided at the school where they studied.
Carter *et al*. 1976 [89]	A study of 114 patients (older than 12 years old) from the Scotland Hemophilia Centre, followed for 3 years. Patients had hemophilia. Outcomes were examined while the integrated care model was in place. Integrated care personnel included a nurse, a senior house officer, a lab technician and a porter. Integrated care included clinic visits and home care.
Chuansumrit *et al*. 1999 [74]	A study of 96 pediatric patients (mean age 9 years old) with congenital bleeding disorders (63 severe and moderate hemophilia A and B, 18 mild hemophilia A, 12 vWD, 3 congenital factor VII deficiency) in Thailand, followed for 20 years. Some patients had inhibitors (17%), HIV (4%) and/or HepC (71%). Outcomes were examined before and after implementing an integrated care model. Integrated care personnel included physicians and health personnel in village health stations, as well as patient assistants in village health stations or district hospitals who delivered primary care and participated in home treatment. Integrated care included home care, education programs, standardized reporting of bleeding episodes and treatment plans, and visits to the integrated care center every 3 to 6 months.
Drake *et al*. 2010 [44]	A study of 7,842 male patients (≥18 years old) with hemophilia A and B from the Universal Data Collection project which includes approximately 130 HTCs in the US., followed between 1998 and 2008. Outcomes were examined while an integrated care model was in place. Integrated care included clinic visits and home care.
Ekert *et al*. 1981 [79, 80]	A study of 7 pediatric patients (5.5-21 years old) with hemophilia A and B at the Royal Children's Hospital in Melbourne, followed for 0.5 to 8 years (median 5 years). Two patients had HepB. Outcomes were examined while the integrated care model was in place. Integrated care personnel included a physician, nurse and social worker. Integrated care included clinic visits and home care.

Heemstra *et al*. 2005 [90]	A study of 17 pediatric patients with severe hemophilia A at the Hospital for Sick Children in Toronto, Canada. Patients were followed from time of diagnosis up to 17-19 years old. Some patients had HIV (47%), HepB (24%) and/or HepC (71%). Outcomes were examined while an integrated care model was in place. Integrated care personnel included a hematologist, nurse, physical therapist, social worker, and external consultants (rheumatologist, orthopaedic surgeon, infectious disease specialist, neurologist, hepatologist, dermatologist, gastroenterologist). Integrated care included clinic visits.
Hilgartner 1977 [81]	A study of 68 patients with hemophilia at New York Hospital-Cornell Medical Center, followed for 3-5 years. Some patients had hepatitis. Outcomes were examined while an integrated care model was in place. Integrated care personnel included hematologist, nurse, physical therapist, social worker, psychiatrist, and other physicians. Integrated care included clinic visits and home care.
Ingram *et al*. 1979 [93]	A study of 28 patients (25% under 19 years old) with severe hemophilia at the Oxford Haemophilia Centre / St. Thomas' Hospital in the UK, followed for 1 year. Some patients had inhibitors (17%), HIV (4%) and/or HepC (71%). Outcomes were examined 3 months before and 9 months after the addition of home care to an integrated care model. Integrated care personnel were not described. Integrated care included clinic visits and home care.
Isarangkura *et al*. 1987 [57]	A study of 10 pediatric patients with hemophilia at the Ramathibodi Hospital, Bangkok, Thailand, followed for 7 months to 7 years (mean 3 years). Some patients had HepB. Outcomes were examined while an integrated care model was in place. Integrated care personnel included local health professionals to assist with infusion, hematologist, nurse/nurse-coordinator, orthopedist, physiatrist, and dentist. Integrated care included clinic visits (1-2 times a year) and home care.
Kennelly *et al*. 1995 [75]	A study of 60 patients (30% <25 years old, 5% ≥45 years old) with hemophilia at the Haemophilia Centre and Haemostasis Unit at the Royal Free Hospital in London, UK, followed for 2 years. Some patients had HIV (58%) and/or HepC (12%). Outcomes were examined while an integrated care model was in place. Integrated care personnel were not described. Integrated care included clinic visits.
Levine and Britten 1973 [55, 77]	A study of 61 pediatric patients with hemophilia at the Blood Coagulation Laboratory of the Hematology Service at the Tufts-New England Medical Center Hospital in Boston. Outcomes were examined while an integrated care model was in place. Integrated care personnel included hematologist, nurse, physical therapist, social worker, orthopaedic surgeon, oral surgeon, and psychologist. Integrated care included home care and clinic visits.

Mahlangu *et al*. 2009 [73]	A study of all patients (59% hemophilia A, 21% vWD, 12% hemophilia B/other) at all 17 HTCs in South Africa, followed for up to 4 years. Some patients had inhibitors (9%). Outcomes were examined while an integrated care model was in place. Integrated care personnel included a hematologist, nurse, physical therapist, social worker, dentist, and geneticist. Integrated care included clinic visits and home care.
Martínez-Murillo *et al*. 2004 [116]	A study of 182 patients (age 0-34 years old) at eight HTCs in Mexico, followed for 1 year. Some patients had inhibitors, HIV, HepB and HepC. Outcomes were examined while an integrated care model was in place. Integrated care personnel included a hematologist and nurse. Integrated care included clinic visits and home care.
McKenzie *et al*. 1974 [92]	A study of 22 patients (>4 years old) with hemophilia A at Vanderbilt University Hospital in Nashville, followed for years. Some patients had inhibitors (5%), HIV (4%) and/or HepC (71%). Outcomes were examined before and after addition of home care to an integrated care model. Integrated care personnel were not described. Integrated care included clinic visits (either at Vanderbilt or through a local physician who works with Vanderbilt closely) and home care.
Niu *et al*. 2014 [47]	A study of 135 patients (of whom 61% were aged 15-64, and 39% were aged 5-14) with hemophilia B (41% severe) enrolled in the HUGS Vb cohort study between June 2009 and April 2013. All patients obtained comprehensive hemophilia care from 10 federally supported HTCs serving 11 geographically diverse US. states. Outcomes were examined while an integrated care model was in place. Integrated care services were those standard to federally supported US. HTCs.
Panicucci *et al*. 1977 [49, 50]	A study of 75 patients with hemophilia at the Centre for the Study and Treatment of Haemophilia and Haemorrhagic Disorders, Pisa, Italy, followed for 2 years. Some patients had inhibitors (17%), HIV (4%) and/or HepC (71%). Outcomes were examined 1 year before and 1 year after the addition of home care to an integrated care model. Integrated care personnel included a hematologist, nurse, physical therapist, orthopaedic surgeon, dentist, genetic counsellor, psychiatrist, and psychologist. Integrated care included clinic visits, telephone contact, 24-hour outpatient care and home care.

Poon *et al*. 2012 [46]	A study of 329 patients (164 adults, 165 children) with hemophilia A enrolled in the HUGS Va cohort study between July 2005 and July 2007. All patients obtained comprehensive hemophilia care from 6 federally supported HTCs serving geographically diverse US states. Outcomes were examined while an integrated care model was in place. Integrated care services were those standard to federally supported US HTCs.
Rabiner *et al*. 1972 [84, 85]	A study of 36 patients (8 months-38 years old) with hemophilia A and B at the Michael Reese Hospital and Medical Center in Chicago, followed for 1 year prior to addition of home care to integrated care and 1 year after addition of home care to integrated care. Outcomes were examined before and after adding home care. Integrated care personnel included hematologists, medical residents and hematology fellows, orthopedic surgeon, psychiatrist, physical therapist and nurse. Integrated care included clinic visits, telephone contact and home care.
Smith *et al*. 1982 [51]	A study of 43 patients (adult and pediatric) with congenital bleeding disorders (hemophilia, and some patients with fXI and fX deficiency) at the Hemophilia Center of Rhode Island/Rhode Island Hospital, followed for 4 years. Some patients had inhibitors (17%), HIV (4%) and/or HepC (71%). Outcomes were examined for 1 year before and 3 years after implementing an integrated care model. Integrated care personnel included hematologist, nurse, full-time nurse coordinator, part-time secretary, physical therapist, social worker, and dentist. Integrated care included clinic visits (at least yearly) and home care.
Smith and Levine 1984 [52]	A study of 4,742 patients with hemophilia A and B at 11 HTCs in the US. Some patients had inhibitors (17%), HIV (4%) and/or HepC (71%). Outcomes were examined for 1 year before and 5 years after implementing an integrated care model. Integrated care personnel included a hematologist, nurse, physical therapist, social worker, internist, pediatrician, orthopaedic surgeon, dentist, a specialized coagulation lab, and a blood bank. Formal linkages with mental health, genetic counseling, rehabilitative services were also in place. Integrated care included clinic visits (at least yearly), home care, a training course in self-therapy (home care), an education program, and an outreach program.

Strawczynski *et al*. 1973 [76]	A comparative study of 40 pediatric patients (2-15 years old) with hemophilia A and B at the Montreal Children's Hospital in Montreal, Canada, followed for 2 years. Some patients had inhibitors (2.5%). Outcomes were examined before and after the addition of home care to an integrated care model. Integrated care personnel included a hematologist, nurse, physical therapist, social worker, pediatrician, orthopedic surgeon, and dental surgeon. Integrated care included all aspects of chronic care, including coordination of consulting services, physiotherapy, immunizations, treatment of infections, and social counselling.
Szucs *et al*. 1998 [82, 83]	A study of 840 patients (35.1 ±14.6 years old) with hemophilia at 16 hemophiliac treatment centers in 10 European countries, enrolled over 6 months. Some patients had HIV (31%), HepB (74%), and/or HepC (46%). Outcomes were examined while an integrated care model was in place. Integrated care personnel were not specified. Integrated care included clinic visits.
Tencer *et al*. 2007 [87]	A study of 61 patients (0-60 years old) with bleeding disorders (hemophilia, vWD) at the Indiana Hemophilia and Thrombosis Center, followed for 1 year prior to initiation of integrated care and 1 year after initiation of integrated care. Some patients had HIV (3%) and/or inhibitors (2.5%). Outcomes were examined before and after implementing an integrated care model. Integrated care personnel included a hematologist, nurse, physical therapist, social worker, a disease management program (DMP) coordinator and specialists in dental hygiene, genetics, nutrition, career counselling, pharmacy, risk reduction, and research. Integrated care included clinic visits and home care.
Weiss *et al*. 1991 [56]	A study of 166 patients (of whom 25% were pediatric) with various congenital bleeding disorders (82.5% hemophilia, 16.9% vWD, 0.6% other) at the Mount Sinai Comprehensive Care Center, followed for 1 year. Outcomes were examined while an integrated care model was in place. Integrated care services included pediatrics, adult medicine, hematology, nursing, physical therapy/rehabilitation, orthopedics, dentistry, genetic counseling, nursing, vocational and educational counseling, as well as psychiatric and social work counseling. Integrated care included clinic visits (1-2 times a year), ongoing contact between visits, and home care in some cases.

HepB: hepatitis B; **HepC:** hepatitis C; **HTC:** Hemophilia Treatment Center; **vWD:** von Willebrand disease; **US**: United States; **UK**: United Kingdom