Improving the Methodology for
Patient-Centred trials in cystic fibrosis

(IMPACT-CF)

Charlie McLeod MBBS (Hons.), FRACP, DTM&H (Dist.), DCH

This thesis is presented for the degree of
Doctor of Philosophy of The University of Western Australia
Postgraduate Research School
Discipline of Medicine

2021
Thesis declaration

I, Charlie McLeod, certify that:

This thesis has been substantially accomplished during enrolment in this degree.

This thesis does not contain material which has been submitted for the award of any other degree or diploma in my name, in any university or other tertiary institution.

In the future, no part of this thesis will be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of The University of Western Australia and where applicable, any partner institution responsible for the joint-award of this degree.

This thesis does not contain any material previously published or written by another person, except where due reference has been made in the text and, where relevant, in the Authorship Declaration that follows.

This thesis does not violate or infringe any copyright, trademark, patent, or other rights whatsoever of any person.

The research involving human data reported in this thesis was assessed and approved by the Child and Adolescent Health Service Human Research Ethics Committee (RGS0000000903). Cross-institutional approval was provided by The University of Western Australia Human Research Ethics Committee (RA/4/20/5548). Written patient consent has been received and archived for the research involving patient data reported in this thesis.

Editorial assistance was provided in preparation for this thesis by Professor Tom Snelling, Professor Steve Webb, A/Professor Richard Norman and A/Professor Christopher Blyth.

This thesis contains published work and/or work prepared for publication, all of which has been co-authored.

Date: 8th of June 2021
Abstract

Cystic fibrosis (CF) is an inherited disease caused by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) protein involving multiple organ systems, for which there is no cure. This protein is involved in the production of mucus, digestive juices and sweat. Pulmonary exacerbations are a hallmark of the disease and are thought to drive lung damage; this leads to progressive lung disease and premature death. Despite a significant investment of effort into research, there is limited high-quality evidence to inform the optimal treatment strategies for these episodes, and consequently the approach to management varies. The lack of high quality evidence is compounded by the inconsistent selection and reporting of outcomes and endpoints in clinical studies and the use of outcomes of uncertain relevance to people affected by CF.

The principal aims of this thesis were: (i) to explore methods for the selection of meaningful outcomes and endpoints in clinical studies in general and (ii) specifically for trials of pulmonary exacerbations in people with CF and (iii) to develop a patient-reported and a proxy carer-reported outcome measure instrument to capture outcomes of importance with the potential for use in future trials of pulmonary exacerbations in adults and children with CF, respectively.

The principal findings of this thesis are:

1. A broad range of outcomes and endpoints have been reported in trials of pulmonary exacerbations in people with CF; many of these outcomes are unlikely to be important to people affected by the disease.
2. There is inconsistency in the application of tests and tools used to capture outcomes in trials among people with CF.
3. Of the many outcomes which may be important to people affected by CF, ten outcomes which are potentially reduced by effective treatment of pulmonary exacerbations are: difficult/painful breathing, sputum production/clearance, fatigue, poor appetite, pain (unrelated to breathing), motivation/demoralisation, fevers/night sweats, treatment burden, inability to meet goals (personal, school, or work) and gastrointestinal symptoms.
4. Of these outcomes, difficult/painful breathing appeared to have the greatest influence on how people affected by CF prioritised their health state preferences regarding treatment for pulmonary exacerbations. Avoidance of gastrointestinal problems also appeared to be highly influential.
This thesis presents two novel weighted outcome measure instruments for use in children (via proxy carer-report) and adolescents and adults with CF (patient-reported), respectively. Ten prioritised outcomes capturing symptoms and functional ability are included in these instruments. The relative importance of these outcomes from the perspective of carers (for the instrument for use in children) and people with CF (for the instrument for use in people older than 13 years old with CF) was based on weights derived from a discrete choice experiment. These instruments are designed to generate a single aggregate score capturing the overall health state of an individual, with scores ranging from 0 (worst health state) to 100 (best health state). The instruments presented in this thesis will be validated in an upcoming clinical trial evaluating different treatment strategies for pulmonary exacerbations in children and adults, BEAT CF. There may be relative advantages afforded by these instruments compared to the alternative patient-reported outcome measures that are currently in use, including the revised cystic fibrosis questionnaire (CFQ-R) and the chronic respiratory infection symptom score (CRISS-CFRSD). The measures presented here contain fewer items than the CFQ-R and CRISS-CFRSD, which is likely to reduce the burden on respondents. We have avoided redundancy in items that capture similar outcomes and little additional information about the overall patient experience. Validation of these instruments will be required however prior to their implementation in clinical research.

Data derived from this thesis will also be used to inform the development of a core outcome set and consensus methods for measurement of outcomes studied in trials of pulmonary exacerbations of CF. This will lead to improved consistency in the selection and reporting of outcomes and facilitate the comparison of results between studies to improve the value of the research that is conducted.
Table of contents

Thesis declaration ............................................................................................................................ iii
Abstract ........................................................................................................................................ iv
Table of contents ........................................................................................................................ vi
Acknowledgements ..................................................................................................................... viii
Declaration ..................................................................................................................................... x
Authorship declaration: co-authored publications ........................................................................ xi
Posters and presentations ........................................................................................................... xviii
List of figures .................................................................................................................................. xx
Appendices ................................................................................................................................... xxi
List of abbreviations .................................................................................................................... xxiii

Chapter 1: Introduction ................................................................................................................ 1
1.1 Thesis overview ...................................................................................................................... 1
1.2 What is CF and what are pulmonary exacerbations? ............................................................ 1
1.3 How common is CF? ............................................................................................................. 2
1.4 Overview of study setting .................................................................................................... 2
1.5 BEAT-CF in context ............................................................................................................ 3
1.6 Thesis in context: why is this work important? .................................................................. 4
1.7 Ethics ................................................................................................................................... 5
1.8 References, appendices and supplementary materials ....................................................... 5
1.9 Chapter summary and thesis aims ...................................................................................... 5

Chapter 2: Estimands framework for designing clinical trials ...................................................... 8
2.1 Chapter summary .................................................................................................................. 8
2.2 Journal article ..................................................................................................................... 9

Chapter 3: Choosing primary endpoints for clinical trials of health care interventions .......... 13
3.1 Chapter summary ................................................................................................................. 13
3.2 Journal article ..................................................................................................................... 14

Chapter 4: Outcomes and endpoints reported in studies of pulmonary exacerbations in people with CF .................................................................................................................. 23
4.1 Chapter summary ................................................................................................................. 23
4.2 Journal article ..................................................................................................................... 24

Chapter 5: Measurement properties of tests and tools used to capture outcomes in CF studies ................................................................. 34
5.1 Chapter summary ................................................................................................................. 34
5.2 Journal article ..................................................................................................................... 35

Chapter 6: Protocol for a discrete choice experiment (DCE) ...................................................... 65
6.1 Chapter summary ................................................................................................................. 65
6.2 Journal article ..................................................................................................................... 66

Chapter 7: A novel method to select outcomes for evaluation in trials and patient-centric outcome measures ................................................................. 72
7.1 Chapter summary ................................................................................................................. 72
7.2 Journal article ..................................................................................................................... 74
Chapter 8: Preferred health outcomes following treatment for pulmonary exacerbations of CF

8.1 Chapter summary..........................................................................................................................98
8.2 Submitted journal article............................................................................................................100

Chapter 9: Conclusions and directions for future research..........................................................121
9.1 Chapter overview .......................................................................................................................121
9.2 Summary of the main findings, strengths and limitations and conclusions from this thesis

9.2.1 Generation of a clinical researcher’s guide for applying an estimands framework when designing clinical trials .................................................................................................................121
9.2.2 Summary of the characteristics, properties and evolution of endpoints used in late phase clinical trials ........................................................................................................................................122
9.2.3 Systematic review of the outcomes and endpoints reported in pulmonary exacerbation trials in people with CF ........................................................................................................................................122
9.2.4 Systematic review of the measurement properties of tests and tools used in trials of CF .123
9.2.5 Development of a novel method to select meaningful outcomes for evaluation in clinical trials ........................................................................................................................................123
9.2.6 Identification of preferred health outcome states resulting from treatment of pulmonary exacerbation episodes and the development of weighted patient and proxy carer-reported outcome measures capturing symptoms and functional impacts in children and adults with CF .124
9.2.7 Capacity building and protocol for the development of a COS for infective exacerbation trials in CF ........................................................................................................................................124
9.3 Where does this thesis sit with relation to the science of clinical trials? ................................125
9.4 Directions for future research ..................................................................................................126

References ........................................................................................................................................147

Appendices .......................................................................................................................................172
Appendix 1: Definitions .....................................................................................................................174
Appendix 2: S1 Initial online health outcomes elicitation survey .....................................................175
Appendix 3: S2 Prioritisation of outcomes survey .............................................................................177
Appendix 4: S3 List of outcomes elicited at workshops and identified in the literature review ...........................................................................................................................................178
Appendix 5 S4: General subnetwork ................................................................................................182
Appendix 6: S5 Gastrointestinal subnetwork .................................................................................183
Appendix 7: S6 Mental health subnetwork .....................................................................................184
Appendix 8: S7 Functional subnetwork ...........................................................................................185
Appendix 9: S1 Example DCE choice set from survey for people with CF .....................................186
Appendix 10: S2 Attributes and levels for the DCE survey for people with CF ...............................187
Appendix 11: S3 Attributes and levels for the DCE survey for carers ............................................188
Appendix 12: S4 DCE survey- adults ............................................................................................189
Appendix 13: S5 DCE survey- adolescents ....................................................................................208
Appendix 14: S6 DCE survey- carers ............................................................................................230
Appendix 15: S7 Retention of participants for DCE survey ................................................................248
Appendix 16: S8 Survey feedback from participants ......................................................................249
Acknowledgements

I have been fortunate to receive support from numerous individuals and groups to enable completion of this work.

First, to Tom Snelling, for providing the vision for this project, for inspiring me to consider the bigger picture, for making the journey fun, for including me in the team, and for your patience, generosity and mentorship. I admire your humility and the way you model leadership, encourage collaboration, promote renewed ways of thinking, and explore novel approaches that challenge the current frontiers of clinical research.

To Steve, for your mentorship; specifically, for teaching me how to craft a narrative, for introducing me to Aristotle’s modes of persuasion (ethos, logos and pathos), and for educating me about other aspects of the dark arts such as the importance of avoiding split-infinitives and compound nouns.

To Richard, for sharing your expertise to facilitate the discrete choice experiment that is central to this work. Thank you for your teaching, your kindness and your patience. The speed with which you provide useful feedback is phenomenal. To Chris, for your clinical leadership, your teaching and for keeping me on track. To Asha, for infecting me with the research bug and for your mentorship and encouragement. To Allison, for your enthusiasm and your willingness to support this work into the post-doctoral phase. To Steven, for your professionalism and for demonstrating compassion and ensuring inclusivity for all participants involved in the Bayesian expert knowledge elicitation workshop.

I am grateful to members of the Sydney Health and Clinical Analytics group and the Telethon Kids Institute Infectious Diseases Implementation Research team for their assistance in facilitating various aspects of this project. Specifically, to Inge Timmerman, for your methodical attention to detail, your organisation, and for bringing colour and humour to the workplace. Thanks also to other team members including Julie Marsh, Sarah Brazier, Nelly Newell, Yue Wu, Kylie Rogers, Danielle Pratt, Tracey Meares, and Grace Curry. I am fortunate to be affiliated with this inspiring group of humans who work cohesively to pursue common goals in the field of infectious diseases research.

I also thank the people affected by cystic fibrosis who contributed their time and subject matter expertise to this project, as well as the consumer and research groups that supported it. Specifically,
to Cystic Fibrosis Australia and their individual state and territory branches. I acknowledge Mitch Messer and Kate Spaapen for helping to recruit participants, and Anne McKenzie for your advice, advocacy and assistance in addressing the ethical challenges we encountered. I am appreciative to Andre Schultz and the Perth Children’s Hospital Respiratory team and the Sir Charles Gairdner Hospital Respiratory team for supporting this work locally, and to Alan Smyth and Sherie Smith for their willingness to collaborate and to promote this work internationally.

I am indebted to the following groups who provided generous financial support to enable this work to progress; the National Medical Health Research Council (NHMRC), the Wesfarmers Centre for Vaccines and Infectious Diseases, the Perth Children’s Hospital Foundation, Cystic Fibrosis Australia and the University of Western Australia. This research was also supported by an Australian Government Research Training Program Scholarship. Without their support, this work would not have been feasible.

Finally, sincere thanks to my family and friends, for supporting me in my endeavour to juggle being a clinician, researcher, partner, friend and mum. To my friends, for their understanding and patience and for kindly providing lifts for my children and gentle reminders about important events (including birthday parties!) not marked on my outlook calendar. To Claire, for the boxing sessions. To Helen, Kev, Sam and Cam, for the hours of babysitting at sub-standard pay. To my parents, for your enduring love, the educational opportunities you have afforded me, and for investing time and energy with our kids to enable our family to balance competing needs. To my sister Bronte, for expanding my mind and for modelling the many ways to demonstrate acts of love. To Lachy, for your surprise packages in the mail. To Jas, for your patience, your love, your spontaneity, for all the little things you do to fill the gaps and damage control when I’m over-committed, and for anchoring me to what’s important. To my three delightful children; Clancy, Archie and Rufus, for making me laugh, for teaching me new tricks, and for reminding me that learning begins by displaying curiosity about the world around you.
Declaration

I dedicate this thesis to individuals and families living with CF, especially those who, despite suffering a significant burden of disease, contributed their time and subject matter expertise to this project in the hope that this research will contribute to improved outcomes for people living with CF in the future.

Most of all, I dedicate this thesis to Jas and my three children; Clancy, Archie and Rufus.
Authorship declaration: co-authored publications

This thesis contains work that has been published and/or prepared for publication.

**Published articles:**

<table>
<thead>
<tr>
<th>Details of the work:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Location in thesis:</strong></td>
<td></td>
</tr>
<tr>
<td>Chapter two</td>
<td></td>
</tr>
<tr>
<td><strong>Student contribution to work:</strong></td>
<td></td>
</tr>
<tr>
<td>Julie Marsh and Tom Snelling broadly conceptualized this piece. I performed the literature review, drafted the original manuscript and constructed Figure 1 and Appendix 1. I drafted the first version of Table 1; the final version incorporates suggestions from Tom Snelling and Julie Marsh. All authors contributed to revisions of the manuscript and approved the final version for publication. I was the first and corresponding author for this manuscript.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Details of the work:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Location in thesis:</strong></td>
<td></td>
</tr>
<tr>
<td>Chapter three</td>
<td></td>
</tr>
<tr>
<td><strong>Student contribution to work:</strong></td>
<td></td>
</tr>
<tr>
<td>Tom Snelling, Steve Webb and I conceptualized this piece. I performed the literature review and drafted the original manuscript. Figure 1 was produced based on a consensus reached between Tom Snelling, Steve Webb and me. All authors contributed to revisions of the manuscript and approved the final version for publication. I was the first and corresponding author for this manuscript.</td>
<td></td>
</tr>
<tr>
<td>Details of the work:</td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Location in thesis:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter four</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Student contribution to work:</th>
</tr>
</thead>
<tbody>
<tr>
<td>I drafted the original protocol for this study and completed the PROSPERO registration. All authors had the opportunity to contribute to the final protocol. Jamie Wood and I performed the search strategy independently. I performed the data collection; this was cross-checked by Jamie Wood. I drafted the primary manuscript. All authors contributed to revisions of the manuscript and approved the final version for publication. I was the first and corresponding author for this manuscript.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Details of the work:</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Location in thesis:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter five</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Student contribution to work:</th>
</tr>
</thead>
<tbody>
<tr>
<td>I drafted the original protocol for this study and completed the registration requirements for PROSPERO. All authors had the opportunity to contribute to the final protocol. Jamie Wood and I performed the search strategy independently. I performed the data collection; this was cross-checked by Jamie Wood. I drafted the primary manuscript. All authors contributed to revisions of the manuscript and approved the final version for publication. I was the first and corresponding author for this manuscript.</td>
</tr>
<tr>
<td>Details of the work:</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td><strong>Location in thesis:</strong></td>
</tr>
<tr>
<td><strong>Student contribution to work:</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Details of the work:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Location in thesis:</strong></td>
<td>Chapter seven</td>
</tr>
<tr>
<td><strong>Student contribution to work:</strong></td>
<td>Tom Snelling was responsible for the overall study concept. I obtained the necessary ethics approvals and was responsible for chairing the workshops conducted in stages one and two involving people with CF and carers of children with CF. Steven Mascaro facilitated the Bayesian Expert Knowledge elicitation workshop. All workshop participants agreed the top 10 priority outcomes and the consensus causal framework. I drafted the primary manuscript. All authors contributed to revisions of the manuscript and approved the final version for publication. I was the first and corresponding author for this manuscript.</td>
</tr>
</tbody>
</table>
Submitted article:

**Details of the work:**

**Location in thesis:**
Chapter eight

**Student contribution to work:**
I obtained the necessary ethics approvals and led the recruitment of participants to this study and data analysis. Statistical support was provided by Richard Norman, who also reviewed the first draft of the manuscript. I drafted the primary manuscript and was the first and corresponding author. All authors contributed to revisions of the manuscript and approved the final version for publication.

For submission within 2 weeks:

**Details of the work:**

**Location in thesis:**
Chapter nine

**Student contribution to work:**
I conceptualised this project and was responsible for drafting the primary manuscript. The study was registered with COMET by Alan Smyth. Primary supervision for this project was provided by Allison Tong. I was the first and corresponding author for this manuscript. All authors contributed to revisions of the manuscript and approved the final version for publication.
Student signature:

Date: 10th of June 2021

Co-authors’ signatures are provided below. This constitutes certification that the students’ statements regarding their contribution to each of the works listed above are correct.

Primary supervisor:

Thomas Snelling
Date: 27.5.21

Coordinating supervisor signature:

Christopher Blyth
Date: 27.5.21

Co-supervisor 1:

Steve Webb
Date: 28.5.21

Co-supervisor 2:

Richard Norman
Date: 25.5.21
Co-authors:

Mitch Messer
Date: 25.5.21

Kate Spaapen
Date: 20.5.21

Matt Stoneham
Date: 28.5.21

Sue Morey
Date: 20.5.21

Siobhain Mulrennan
Date: 21.5.21

Andre Schultz
Date: 20.5.21

Steven Mascaro
Date: 25.5.21
Posters and presentations


McLeod C. Improving the methodology for patient-centred trials in cystic fibrosis. Wesfarmers Retreat. November 2018

Upcoming presentations:


Abstract submitted for consideration:

List of figures

Chapter one

*Figure 1* Map of Perth, Australia

List of supplementary tables and appendices
Appendices

Chapter 2

Appendix 1: Definitions

Chapter 3 (available electronically)

Appendix A: Search strategy strand 1 (Medline)
Appendix B: Search strategy strand 2 (Medline)

Chapter 4 (available electronically)

Appendix 1: Abbreviations and their meanings
S1: Search strategy (Medline)
S2: Search strategy (Embase)
S3: Published studies that met inclusion criteria
S4: Strengths and limitations of identified outcomes

Chapter 5 (available electronically)

Appendix 1: COSMIN Definitions for measurement properties
Appendix 2: Abbreviations and their meanings
S1: Search strategy (Medline)
S2: Search strategy (Embase)
S3: Studies that met inclusion criteria
S5: Characteristics of tests and tools
S6: Measurement properties of tests and tools

Chapter 6 (available electronically)

Appendix 1: Online CF-related health outcomes survey
Chapter 7

Appendix 2: S1 Initial online outcomes elicitation survey
Appendix 3: S2 Prioritisation of outcomes survey
Appendix 4: S3 List of outcomes elicited at workshops and identified in the literature review
Appendix 5: S4 General subnetwork
Appendix 6: S5 Gastrointestinal subnetwork
Appendix 7: S6 Mental health subnetwork
Appendix 8: S7 Functional subnetwork

Chapter 8

Appendix 9: S1 Example DCE choice set from survey for people with CF
Appendix 10: S2 Attributes and levels for the DCE survey for people with CF
Appendix 11: S3 Attributes and levels for the DCE survey for carers
Appendix 12: S4 DCE survey- people with CF
Appendix 13: S5 DCE survey- adolescents with CF
Appendix 14: S6 DCE survey- carers
Appendix 15: S7 Retention of participants for DCE survey
Appendix 16: S8 Survey feedback from participants
# List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>APT</td>
<td>Adaptive platform trial(s)</td>
</tr>
<tr>
<td>BEAT CF</td>
<td>Bayesian evidence adaptive treatment of pulmonary exacerbations of cystic fibrosis</td>
</tr>
<tr>
<td>CF</td>
<td>Cystic fibrosis</td>
</tr>
<tr>
<td>CFTR</td>
<td>Cystic fibrosis transmembrane regulator protein</td>
</tr>
<tr>
<td>COS-PEX</td>
<td>Core outcome set for evaluation in pulmonary exacerbation studies of cystic fibrosis</td>
</tr>
<tr>
<td>DCE</td>
<td>Discrete choice experiment</td>
</tr>
<tr>
<td>MAOI</td>
<td>Multi-attribute outcome instrument</td>
</tr>
<tr>
<td>ppFEV1</td>
<td>Partial pressure forced expiratory volume in one-second</td>
</tr>
<tr>
<td>QoL</td>
<td>Quality of life</td>
</tr>
<tr>
<td>U.S.</td>
<td>United States</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

1.1 Thesis overview

This chapter will explain the rationale for this thesis and introduce the major themes. The genesis of this thesis was to inform the design of BEAT CF; a multi-site adaptive platform trial evaluating the efficacy and safety of treatment strategies for pulmonary exacerbations in people with cystic fibrosis (CF). Specifically, we aimed to develop a multi-attribute outcome instrument (MAOI) to capture a number of patient-centred outcomes relevant to pulmonary exacerbations; one for use in children and another for use in adolescents and adults. It was envisaged that a change in the MAOI score from baseline could be implemented as a single patient-centred endpoint to evaluate interventions studied in future trials in people with CF like BEAT CF.

This thesis comprises two parts. Part 1 focuses on methods used to select outcomes and their corresponding endpoints for clinical trials. We begin by evaluating how and why outcomes and their corresponding endpoints are chosen; Chapter 2 explains the merits of the estimands framework when designing clinical trials and Chapter 3 reviews factors that should inform the selection of endpoints. Part 2 begins with Chapters 4 and 5, which report the findings of two systematic reviews; the first evaluates the outcomes and endpoints reported in studies of pulmonary exacerbations in people with CF, and the second describes the measurement properties of tests and tools used to capture outcomes in studies in people with CF. In Chapter 6, I present a protocol for the projects central to this thesis; including (i) the development of a novel approach to selecting outcomes for clinical trials, applied to trials among people with CF and (ii) a discrete choice experiment (DCE) to understand the relative importance of each outcome among people affected by CF, and for deriving weights for each outcome when combined in a MAOI. Chapter 7 reports the ten prioritized outcomes selected for inclusion in the weighted outcome measure instruments and Chapter 8 presents the results of the DCE. Chapter 9 concludes the thesis with a discussion of the results and proposes directions for future research. A protocol for the development of a core outcome set for evaluation in pulmonary exacerbation trials in CF (COS-PEx) is included here; results of this post-doctoral endeavor are expected to follow within 12-months of submission of this thesis.

1.2 What is CF and what are pulmonary exacerbations?

CF is an inherited, life-limiting disease affecting multiple organs, predominantly the lungs
(Knowles, 2017), for which there is no cure. The median age at death is 35 years for people living with disease in Australia, and the major pathway to premature death is progressive lung disease leading to respiratory failure (Ruseckaite, 2019).

Pulmonary exacerbations are a hallmark of CF and are thought to drive progressive lung damage; 25% of people don’t regain their baseline lung function after these episodes (Sanders, 2010). There is no consensus definition for pulmonary exacerbations, although these episodes generally involve a deterioration in lung function and new or worsening respiratory and systemic symptoms (Sanders, 2010). Management of these episodes is complex, invasive, and often requires prolonged hospitalisation (Hurley, 2015). While no agreement exists regarding the best treatment strategies, management generally involves a combination of antimicrobial and airway clearance therapies (including physiotherapy with or without muco-active agents) and optimisation of nutrition (Hurley 2015; Smyth A, 2017; Abbott, 2019, Bhatt, 2019; Jain, 2018; Waters, 2015).

1.3 How common is CF?

The incidence of CF varies worldwide; the overall prevalence is estimated to be between 70,000 – 100,000 (CF Worldwide, 2020). In Australia, one in 2,500 babies are born with CF (between 70 to 100 new diagnoses per annum) (CF Registry report, 2019); this is comparable to one in 2,000-3,000 babies born in Europe and one in 3,500 babies born in the United States of America (CF Worldwide, 2020). In Ireland, one in 19 people carry a defective CF gene. Tasmania has the second highest rate of CF gene carriage in the world behind Ireland (CF Ireland, 2019), with one in 20 people carrying a defective gene (CF Australia, 2019).

1.4 Overview of study setting

The projects central to this thesis were conducted in partnership with people affected by disease. Recruitment predominantly occurred within Australia through: (i) inpatient and outpatient facilities at the Perth Children’s Hospital (PCH), Western Australia’s only tertiary pediatric facility, (ii) the outpatient service at Sir Charles Gairdner Hospital, a large tertiary adult facility also in Perth, (iii) via snowballing strategies using professional and research networks including CF Australia, the Telethon Kids Institute CONNECT network and the Australian Society of Infectious Diseases, and (iv) CF consumer and research networks within Australia and overseas, including via communiques and electronic media (such as Facebook and Twitter).
All consumer engagement activities involving people with CF were conducted virtually owing the infection control restrictions which preclude mixing of this patient population.

![Perth, Australia (red pin) [source: google maps]](image)

**Figure 1** Perth, Australia (red pin) [source: google maps].

### 1.5 BEAT-CF in context

There is a lack of consensus regarding the optimal management for pulmonary exacerbations of CF. Contributing factors are low patient numbers, weak clinical research infrastructure, limited capacity, lack of collaborative will and financial support for robust clinical trials, and the selection of outcomes of little meaning to people living with disease (McLeod, 2020).

Randomised controlled trials are rightfully considered the gold standard approach for generating evidence to inform clinical practice and policy (Moher, 2015). However, conventional trials can be burdensome, inefficient and limited to addressing only one question at a time (Kramer, 2006).

With advances in statistical modelling, adaptive platform trials (APTs) are now feasible and this approach offers potential benefits of flexibility and efficiency over conventional trials in select circumstances (Angus, 2019, Kramer 2006)). An APT studies multiple interventions for a single disease or condition within a single trial platform. Intervention arms may be added or removed depending on pre-specified rules that are determined *a priori* (Angus, 2019).
BEAT-CF is an adaptive platform trial nested within a prospective multi-site cohort of people with CF receiving treatment for infective exacerbations. BEAT-CF trial has been designed, and will be conducted in partnership with, key stakeholders including people with lived experience of CF. The trial will be embedded in routine clinical care and aims to create a ‘learning health system’ approach to CF care by ensuring the rapid implementation of new knowledge into clinical practice (Angus, 2019). Arguably, APTs may be more ethically acceptable than conventional fixed-design trials because the design features allow new knowledge to be generated efficiently and because participants stand to benefit directly from accumulating trial data. APTs may also be more cost-effective than conventional trial designs (Pallman, 2018).

1.6 Thesis in context: why is this work important?

Historically, the selection of outcomes, endpoints and the tools for assessing outcomes in trials has largely been the prerogative of researchers; people living with disease have had little or no role in this process (Moher, 2016; Al-Shahi, 2014). Since many biological outcomes may be objectively and easily captured (for example lung function measured as the percentage predicted forced expiratory volume in one-second [ppFEV1]), many researchers will be motivated to use these rather than outcomes of direct relevance to people affected by disease – such as symptom resolution, quality of life, ability to return to school or work and the burden of treatment or disease.

For an outcome to be meaningful, it should arguably capture how a person feels, functions or survives either directly or else it should be a reliable surrogate of those outcomes (US FDA, 2018), and should be acknowledged as important to people affected by disease. Failure to select meaningful outcomes for evaluation in clinical trials may diminish the value of trials that seek to improve the lives of those affected by disease and results in wasted research investment (US FDA, 2018; Chalmers, 2014; Ionnidis, 2014).

While involving consumers in the process of outcome selection for clinical research is increasingly recognised as important (U.S FDA, 2018), this is only occasionally done. This thesis has engaged people affected by CF to determine which outcomes they consider to be meaningful to them, and uses a novel approach to select a group of prioritized outcomes for inclusion in two MAOI’s (one for use in adolescents and adults with CF and one for use in children with CF); the intention is the MAOI will be applied as a summary endpoint for evaluating alternative treatment strategies in the BEAT CF trial. This thesis also involves necessary steps towards the development of a core outcome set (COS) for pulmonary exacerbation trials of CF. A COS is a collection of outcomes
derived by broad stakeholder consensus that should be measured and reported in all trials for a specific condition (Williamson, 2012). COSs are important to encourage standardisation in the selection and reporting of outcomes for clinical trials in order to improve the value of research.

1.7 Ethics

This study was approved by the Child and Adolescent Human Research Ethics Committee (RGS0000000903). Cross-institutional approval was also provided by the University of Western Australia Human Research Ethics Committee (RA/4/20/5548). Ethics approval was current throughout the period of the study.

1.8 References, appendices and supplementary materials

References are included at the end of published and submitted articles as they appear in the thesis. An electronic link to appendices and supplementary materials is provided, where possible, or otherwise at the end of the thesis. A consolidated list of references including those cited in the linking text is also supplied at the end of the thesis. References are cited throughout the linking text as (surname of first author, year of publication).

1.9 Chapter summary and thesis aims

Cystic fibrosis is a rare, inheritable disease which is characterized by intermittent pulmonary exacerbations which drive progressive lung damage; this in turn leads to premature death predominantly due to respiratory failure.

There is limited high quality evidence to guide management for pulmonary exacerbations in people with CF. In part, this may be because outcomes selected for evaluation in clinical trials may not be meaningful to those living with disease. This thesis investigates which outcomes are meaningful to people affected by CF and presents a novel method for developing weighted outcome measures as a patient-centric tool for evaluating the impact of treatment strategies studied in clinical trials of pulmonary exacerbations in children and adults with CF.

The specific aims of this thesis are:
1. To describe the application of the estimands framework to the design of clinical trials.

2. To review the considerations that should inform the selection of outcomes and endpoints in clinical trials.

3. To review the outcomes and endpoints which have been reported in trials for pulmonary exacerbations in CF.

4. To summarise the measurement properties of the tests and tools used to measure outcomes reported in trials in CF.

5. To develop an approach for selecting outcomes for clinical trials based on (i) their importance to people with lived experience of the disease, (ii) their causal relation to the underlying pathophysiological processes and (iii) the likelihood that they will be impacted by the intervention under study, and to apply this to select outcomes for trials in pulmonary exacerbations of CF.

6. To develop MAOI’s for evaluating patient-reported and proxy carer-reported outcomes in trials evaluating treatment strategies for pulmonary exacerbations in adults and adolescents with CF and children with CF, respectively.
Part I
Chapter 2: Estimands framework for designing clinical trials

2.1 Chapter summary

This chapter describes the estimands framework, and its application to clinical trials to ensure that the design, conduct, and analyses align with the trial objectives. While much of the discussion of the estimands framework has been with respect to regulatory trials, its broader application will help ensure that results generated from all trials are meaningful and can be translated into practice and policy, and thereby reduce research waste (Akacha, 2017; Chalmers, 2014; Ioannidis, 2014).

Herein I describe the five attributes of an estimand; the target population, a description of the intervention, the endpoint, population-level summary, and adjustments for intercurrent events and illustrate these concepts with examples drawn from trials of interventions for people with pulmonary exacerbations of CF.

2.2 Journal article


Appendix 1: Definitions
A clinical researcher’s guide for estimands: how they relate to outcomes, endpoints and measures of treatment effect in clinical trials

Charlie McLeod¹-³, Julie Marsh¹, Allison Tong⁴, Christopher C Blyth¹-³,⁵, Richard Norman⁶, Steve Webb⁷,⁸ and Thomas L Snelling¹⁹

¹Westmead Centre for Vaccines and Infectious Diseases, Telethon Kids Institute, University of Western Australia, Australia
²School of Medicine, University of Western Australia, Australia
³Infectious Diseases Department, Perth Children’s Hospital, Australia
⁴Sydney School of Public Health, Faculty of Medicine and Health, University of Sydney, Australia
⁵Pathwest Laboratory Medicine WA, QEII Medical Centre, Australia
⁶School of Public Health, Curtin University, Australia
⁷St John of God Hospital, Australia
⁸School of Population Health and Preventive Medicine, Monash University, Australia
⁹Menzies School of Health Research, Australia

Clinical trials are conducted to inform practice and policy by generating evidence of the efficacy and safety of interventions [1,2]. Most focus on creating generalizable knowledge by demonstrating treatment effects and by quantifying their size. Key steps in planning involve defining the research question, determining the data required to answer the question and how this will be analysed, and translating the evidence generated by implementing changes to practice and policy where appropriate [3,4]. However, greater transparency is required on how trial objectives are translated into treatment effects, which requires a clear dialogue between the disciplines involved in designing, executing and reporting a trial.

Designing trials in a way that ensures the results are meaningful and can be readily translated into practice is critical for reducing research waste [5]. However, the challenges are not always acknowledged. Firstly, there is inconsistency in the nomenclature used to describe key clinical trial design features (Appendix 1) [6-8]. Secondly, trial objectives don’t always align with statistical approaches [9]. This can occur when adjustments are needed to account for missing data or for events that occur following randomization (known as intercurrent events) that preclude the observation of the outcome or affect its interpretation [10]. In 2010, the National Research Council of the United States of America drew attention to these issues and proposed a new approach for designing trials based on defining the specific target for estimation, the estimand [11]. An estimand consists of five attributes that each address a trial objective: target population, descriptions of interventions, endpoint, population-level summary and adjustments for intercurrent events [12]. A framework has been established around this concept to assist in planning trials to ensure that the design, conduct and analysis aligns with its objectives. While this framework has been well described in the statistical literature, it has not been uniformly adopted by the clinical research community. It is hoped that this approach will improve the transparency and value of the research that is conducted [9,13]. A common misconception is that estimands are a statistical concern as they have been defined in an addendum to the ICH E9 statistical guidance for clinical trials [13].

Here we review the concepts of outcomes, endpoints and population level summaries of treatment effects, and introduce the concept of estimands for clinical researchers. We aim to (i) provide clarification of these concepts to ensure they are not inadvertently conflated, (ii) encourage consistent use of terminology to avoid confusion and (iii) promote the explicit use of estimands when designing clinical trials in multidisciplinary teams and communicating results to policy-makers and consumers [14]. We illustrate the concepts with examples drawn from trials of interventions for people with pulmonary exacerbations of cystic fibrosis (CF).

Outcomes

Outcomes are the characteristics or biological processes that are potentially affected by an intervention, and a beneficial effect on the primary outcome (an increase or decrease, promotion or prevention, depending on the outcome) is generally the primary intention of our intervention. A range of distinct outcomes occur among trial participants [7]; some occur because of or are modified by an intervention (or its absence), while many more occur unaffected by or despite an intervention [15]. Outcome(s) chosen for evaluation in trials should address the trial objective(s) by capturing the range of benefits and costs attributable to an intervention. To do this, chosen outcomes should be meaningful, such that they should arguably reflect (directly or indirectly) how a person feels (such as shortness of breath or chest pain), functions (such as capacity to exercise or climb stairs) or survives, and should be acknowledged as being important to patients [2,8,16,17].

The International Society for Pharmacoeconomics and Outcomes Research taskforce (ISPOR) categorise outcomes as clinical or non-clinical [18]. Clinical outcomes are those that may be influenced by human choice, judgement or motivation [18], being either: (i) reported by clinicians based on clinical events (e.g. pulmonary exacerbation or lung transplant), (ii) assessed by standardised performance measures (e.g. forced expiratory volume in 1-second [FEV1])) (iii) reported by patients (e.g. perceived quality of life), or (iv) observer-reported (e.g. parent-reported treatment adherence) [18,19]. While non-clinical outcomes such
as biomarkers (e.g. c-reactive protein) or pathological changes (e.g. presence of bronchiectasis) don’t directly capture how a person feels, functions or survives, they may still be informative for clinical decision making. They may provide evidence of the presence or severity of underlying pathophysiological processes, without capturing the clinical manifestations or consequences of those processes per se. Evaluation of these outcomes may help to understand the effect of treatment on the causal processes underlying disease [17]; to the extent that non-clinical outcomes are associated with, or predictive of, how someone feels, functions or survives, they may still be informative for clinical decision making, or for determining whether an intervention should undergo further evaluation, for example in larger pre-licensure trials.

**Endpoints**

Whereas the outcome is the clinical, biologic or pathophysiological characteristic or process of interest, the endpoint defines the specific analysis parameter captured as evidence of that outcome in a clinical trial (e.g. the change in the percentage predicted FEV1 from baseline to day 14) [20,21]. An outcome may be captured using different endpoints. The ability of an endpoint to capture the outcome of interest depends on the extent to which it measures the outcome with sufficient reliability, precision and validity [16].

Endpoints may capture the presence of an outcome (state or characteristic of a person) at a specific time (e.g. survival at 12 months) or a change in a persons’ health state or characteristic over a defined period (e.g. the change in patient-reported breathlessness from the start to end of treatment) [17].

Endpoints may be summarised by their central tendency and distribution (e.g. mean and standard deviation) for continuous measures, or their frequency (e.g. as risk or rates) for discrete events for each treatment group. Importantly, this group level summary of the endpoint should not be confused with the endpoint itself, or with the population level summary of the treatment effect.

**Population level summary of treatment effects**

Because the effect of a treatment can rarely be known or quantified for an individual, it is more usual for the treatment effect to be inferred by contrasting or statistically modeling the summary of the endpoint (central tendency or frequency) at the group level between interventions. If the endpoint selected for a trial is an event (e.g. hospitalisation), the effect of an intervention on that endpoint may be quantified as a difference or ratio of the risk or rate of the endpoint between the treatment arms (e.g. risk difference or rate ratio of hospitalisation); it is also possible to use the risk difference to express the treatment effect in terms of the number of patients needed to treat, on average, to prevent one event [22]. If a continuous endpoint (e.g. survival over 10 years) is selected, the treatment effect may be quantified as the difference in the group means or medians [23].

The measure of a treatment effect may also be a parameter derived from a statistical model (e.g. the odds ratio estimated from logistic regression or the hazard ratio estimated from a Cox proportional hazard model). An appropriate model must be selected for use, because if the statistical model does not reflect the data generating process, then interpretation of the empirical summary may be difficult. For example, a single hazard ratio may not appropriately summarise the effect of antibiotics on increasing the average time-to-transplantation if resistance causes the effect to wane over time, such that the odds ratio is not constant [24]. Further, odds and hazards ratios are not well understood and are often conflated with more intuitive measures like risk or rate ratios, but they are distinct measures of effect and may not always concord [25]. While risk and rate ratios are intuitive, odds ratios and hazard ratios are arguably less well understood. In all cases, both the point estimate of the treatment effect (the estimate best supported by the data) and an indication of the uncertainty of this estimate or range of plausible estimates of the effect (e.g. a 95% confidence or credible interval), should be reported.

![Figure 1. Five key attributes of an estimand](image)
Estimand(s)

Estimand(s) define precisely what treatment effect you are trying to estimate in a trial [9]; these need to be specified a priori. The five key attributes of an estimand are depicted in figure 1 [12,13].

The population of interest is those individuals targeted by the study question (e.g. children with CF hospitalised for pulmonary exacerbations) and is defined by specification of the inclusion (e.g. 0-17 years old) and exclusion criteria (e.g. lung transplant recipients). The treatment description refers to the exact intervention under evaluation (e.g. a 14-day intravenous course of ceftazidime plus tobramycin) and the comparator treatment (e.g. standard of care). The endpoint (e.g. the absolute change in percentage of predicted forced expiratory volume in 1-second between baseline and day 14) must capture the outcome of interest (e.g. lung function) and align with the research question (e.g. what is the expected improvement in lung function if I give a child the investigational treatment rather than standard therapy?).

Intercurrent events are those that occur after enrolment (or randomisation) and before ascertainment of the endpoint which may preclude the observation of an outcome or affect its interpretation (e.g. death, discontinuation of therapy due to toxicity or adverse events or use of rescue medications) [13]. For example, if treated patients are more likely to survive, their length of hospitalization may increase. Combining outcomes like hospitalization and death into ‘hospital-free survival’ may help to circumvent these problems [20]. Many analytic strategies are proposed for dealing with intercurrent events [12], but the impact of the chosen strategy on the study objective should also be considered. The ICH guideline on estimands outlines five common strategies (treatment policy, composite, hypothetical, principal stratum, while on treatment) but researchers are not restricted to these strategies, which should be specified in the protocol for each type of intercurrent event [12]. Intercurrent events should be distinguished from events that result in missing data (e.g. study discontinuation due to loss to follow-up or incomplete data collection), as the approaches used to handle these scenarios will vary (e.g. it may be stipulated that participants who receive adjunctive or rescue therapies will have their results included in the analysis) [12,14,26]. The population level parameter describes how the effect of treatment will be quantified, usually as a comparison between treatment arms (e.g. the point estimate and 95% confidence interval of the difference in the mean change in percentage predicted FEV1 between active and placebo groups).

When formulating estimand(s), it may be necessary to specify one or more sensitivity analyses to explore the impact of any assumptions about the data on the estimate of the treatment effect (population level summary). For example, differential use of adjunctive or rescue therapies may mask the benefit of a treatment and it may be impossible to know what would have occurred if those participants had not received them. A sensitivity analysis may assume that those who receive such therapies would have had a worse outcome than that observed (e.g. a sensitivity analysis may assume that participants who receive adjunctive therapy would have had no improvement in FEV1 if they had not received the therapy) [13].

In certain situations, it may be desirable to design a trial to identify multiple similarly effective treatments, rather than the single most effective treatment (e.g. antibiotics for the treatment of pulmonary exacerbations). This objective may be efficiently addressed using a platform adaptive trial design incorporating response-adaptive randomisation, but it may be necessary to compromise the estimand in order to achieve this. Imposing a decision rule in an adaptive trial that will remove a treatment option can result in an intrinsic but small bias if there are insufficient allocations to this treatment option to ensure regression to the true group level summary [27]. It may be that researchers are willing to allow a minimal bias in order to achieve optimisation of outcomes for the population under study, depending on the purpose of the study.

For illustration, table 1 provides further examples of estimands for trials of pulmonary exacerbations in people with cystic fibrosis.

Conclusions

Clinician researchers should be encouraged to use estimands to facilitate a clear dialogue between the disciplines when designing clinical trials. An estimand must align with the objectives of the trial, the study design, the data collected and the methods for analysis and reporting. Use of this framework is likely to improve the transparency and value of the research that is conducted. Depending on the purpose of the trial, it may be necessary to minimally compromise the estimate of a treatment effect in order to achieve optimisation of resources, results and outcomes for the population under study.

<table>
<thead>
<tr>
<th>Estimand(s)</th>
<th>Estimates precise what treatment effect you are trying to estimate in a trial [9]; these need to be specified a priori. The five key attributes of an estimand are depicted in figure 1 [12,13].</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population of interest</td>
<td>The population of interest is those individuals targeted by the study question (e.g. children with CF hospitalised for pulmonary exacerbations) and is defined by specification of the inclusion (e.g. 0-17 years old) and exclusion criteria (e.g. lung transplant recipients).</td>
</tr>
<tr>
<td>Treatment description</td>
<td>The treatment description refers to the exact intervention under evaluation (e.g. a 14-day intravenous course of ceftazidime plus tobramycin) and the comparator treatment (e.g. standard of care).</td>
</tr>
<tr>
<td>Endpoint</td>
<td>The endpoint (e.g. the absolute change in percentage of predicted forced expiratory volume in 1-second between baseline and day 14) must capture the outcome of interest (e.g. lung function) and align with the research question (e.g. what is the expected improvement in lung function if I give a child the investigational treatment rather than standard therapy?).</td>
</tr>
<tr>
<td>Intercurrent events</td>
<td>Intercurrent events are those that occur after enrolment (or randomisation) and before ascertainment of the endpoint which may preclude the observation of an outcome or affect its interpretation (e.g. death, discontinuation of therapy due to toxicity or adverse events or use of rescue medications) [13].</td>
</tr>
<tr>
<td>Population parameter</td>
<td>The population level parameter describes how the effect of treatment will be quantified, usually as a comparison between treatment arms (e.g. the point estimate and 95% confidence interval of the difference in the mean change in percentage predicted FEV1 between active and placebo groups).</td>
</tr>
</tbody>
</table>

When formulating estimand(s), it may be necessary to specify one or more sensitivity analyses to explore the impact of any assumptions about the data on the estimate of the treatment effect (population level summary). For example, differential use of adjunctive or rescue therapies may mask the benefit of a treatment and it may be impossible to know what would have occurred if those participants had not received them. A sensitivity analysis may assume that those who receive such therapies would have had a worse outcome than that observed (e.g. a sensitivity analysis may assume that participants who receive adjunctive therapy would have had no improvement in FEV1 if they had not received the therapy) [13].

In certain situations, it may be desirable to design a trial to identify multiple similarly effective treatments, rather than the single most effective treatment (e.g. antibiotics for the treatment of pulmonary exacerbations). This objective may be efficiently addressed using a platform adaptive trial design incorporating response-adaptive randomisation, but it may be necessary to compromise the estimand in order to achieve this. Imposing a decision rule in an adaptive trial that will remove a treatment option can result in an intrinsic but small bias if there are insufficient allocations to this treatment option to ensure regression to the true group level summary [27].

It may be that researchers are willing to allow a minimal bias in order to achieve optimisation of outcomes for the population under study, depending on the purpose of the study.

For illustration, table 1 provides further examples of estimands for trials of pulmonary exacerbations in people with cystic fibrosis.

Conclusions

Clinician researchers should be encouraged to use estimands to facilitate a clear dialogue between the disciplines when designing clinical trials. An estimand must align with the objectives of the trial, the study design, the data collected and the methods for analysis and reporting. Use of this framework is likely to improve the transparency and value of the research that is conducted. Depending on the purpose of the trial, it may be necessary to minimally compromise the estimate of a treatment effect in order to achieve optimisation of resources, results and outcomes for the population under study.
Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgements

No specific funding was received for this work. CM is supported by an NHMRC post-graduate scholarship (GNT1150996), and grants from the Wesfarmers Centre (top-up grant) and the Perth Children’s Hospital Foundation (9772 & 9757). TS is supported by a Career Development Fellowship from the National Health and Medical Research Council (GNT1111657). CB is also supported by the National Health and Medical Research Council (GNT1111596/1173163).

References

Chapter 3: Choosing primary endpoints for clinical trials of health care interventions

3.1 Chapter summary

This chapter reviews the evolution, the range and the relative strengths and weaknesses of endpoints used in late phase trials; these are trials that are designed to generate evidence to inform clinical practice and policy. We define clinical and non-clinical endpoints and distinguish those that are likely to be meaningful. Arguably, for an endpoint to be meaningful, it should capture how a person feels, functions or survives and be acknowledged as important to people living with disease or be a validated surrogate of these outcomes (U.S FDA, 2018).

This review is intended to serve as a reference for assisting researchers when choosing primary endpoints and for the end-users of clinical trial data tasked with translating this evidence into clinical practice and/or policy.

3.2 Journal article


Appendix A: Search strategy strand 1 (Medline)

Appendix B: Search strategy strand 2 (Medline)

Appendices are available electronically here.
Choosing primary endpoints for clinical trials of health care interventions

Steve Webbe, Thomas L. Snelling

Received 1 June 2019; Received in revised form 29 October 2019; Accepted 9 November 2019

The purpose of late phase clinical trials is to generate evidence of sufficient validity and generalisability to be translated into practice and policy to improve health outcomes. It is therefore crucial that the chosen endpoints are meaningful to the clinicians, patients and policymakers that are the end-users of evidence generated by these trials. The choice of endpoints may be improved by understanding their characteristics and properties. This narrative review describes the evolution, range and relative strengths and weaknesses of endpoints used in late phase trials. It is intended to serve as a reference to assist those designing trials when choosing primary endpoint(s), and for the end-users charged with interpreting these trials to inform practice and policy.

1. Introduction: Purpose of clinical trials and why endpoint selection is important

The purpose of late phase trials is to generate evidence to guide decision-making in clinical practice and in policy. In this regard, clinicians, patients, and policymakers are all end-users of clinical trial evidence. Randomised clinical trials represent a gold standard for generating evidence, as they are the least biased way of measuring and comparing treatment effects [1].

Many outcomes occur among trial participants [2]; some outcomes occur because of an intervention or because of the absence of one, some outcomes may be modified by an intervention (for example, time to event or severity), while many more outcomes occur unaffected by an intervention. Outcome(s) selected for evaluation must address the trial objective(s) and should be acknowledged as meaningful to end-users. For an outcome to be meaningful, it should reflect or describe how a person feels, functions and survives [3]. Endpoints are the specific measures of these outcomes [2]. If end-users are going to make decisions based on measured differences in one or more endpoints between treatment groups, they must understand what those differences are; but endpoints have properties and characteristics that have strengths and limitations that are critical to their interpretation.

It is a responsibility of those who design and conduct trials to choose endpoints which will influence decision-making by clinicians and policymakers. Endpoint selection is a complex process. End-users bring differing needs and perspectives. Poor selection of endpoints makes interpretation and implementation of findings difficult or impossible, limits evidence synthesis, and thereby diminishes the value of the research, resulting in wasted use of resources [4].

A single endpoint may not capture the important effects of an intervention to the satisfaction of all end-user groups, so multiple endpoints are usually selected, which are categorized as primary, secondary or tertiary. Primary endpoint(s) are typically efficacy measures that address the main research question [3]. Secondary endpoints are generally not sufficient to influence decision-making alone, but may support the claim of efficacy by demonstrating additional effects or by supporting a causal mechanism [2]. If tertiary endpoints are nominated, they typically capture outcomes that occur less frequently or which may
be useful for exploring novel hypotheses [3].

The primary aim of this review is to summarise the range of clinical and non-clinical endpoints used in late phase trials and their relative strengths and weaknesses. The secondary aims are to describe their evolution and consider which characteristics of endpoints are valuable for evaluating treatment effects. This review is intended to serve as a reference to assist researchers when choosing primary endpoints, and for the end-users of clinical trial data tasked with translating this evidence into clinical practice or policy. Early phase trials may have a more proximal aim such as establishing proof-of-principle, trial feasibility, or assessing the mechanistic effects of an intervention; this review does not discuss endpoints relevant to these types of trials. Further, whilst we recognise that statistical and regulatory considerations are also important factors weighing into overall endpoint selection, a detailed analysis of these topics is beyond the scope of this review.

2. Methods

We developed a two-strand search method to address our research questions, incorporating the following search terms using a Boolean strategy, including papers published up to October 2018: “Endpoint determination”, “surrogate, biomarker, combination, individuality”,” multiple or composite,” “end-point,” “Outcome Assessment (Health Care),” “Research Design” and “Clinical trials.” This search was executed in Medline (Medline and Epub Ahead of Print, In-Process and other non-indexed citations 1946-) and Embase (Embase & Classic 1947-) and limited to articles written in English. Registration guidelines issued by the Food and Drug Administration and European Medicine Association (EMA) were also examined using the same keywords. Additional articles were identified through citation review of selected articles and some clinical examples were drawn from the authors’ experience. Our full search strategy (including additional limits) is detailed in Appendix A and B. The search was performed by a single reviewer (CM) and findings are reported by narrative synthesis.

3. Results

3.1. Classification of endpoints

Endpoints for late phase trials can be broadly classified as either clinical or non-clinical (see Fig. 1) [3,5].

Clinically meaningful endpoints relate to outcomes which capture how a person feels, functions or survives [3]. These endpoints may be measured objectively or subjectively, and are either (i) reported by clinicians (ClinRO), which involves judgement or interpretation of clinical signs or events (such as stroke, myocardial infarct or cancer remission), (ii) assessed by standardised performance measures (6-min walk test), (iii) patient-reported (PRO), which are directly reported by patients (such as self-reported symptoms or function, or a measure of perceived quality of life) or (iv) observer-reported (ObsRO), such as a parent log of seizure activity in a child [5].

Non-clinical endpoints, including biomarkers, do not relate directly to how a person feels, functions or survives, but are instead objectively measured indicators of a biological or pathogenic process, for example a pharmacological response to a treatment intervention. Biomarkers may include blood tests (for example laboratory measures such as troponin and haemoglobin concentration or serological assays), tissue/fluid analyses (for example histopathological results), imaging results, or physiological measures (for example blood pressure) which are used for diagnostic, prognostic, monitoring (including safety) or predictive purposes [2].

Some endpoints may be clinically important even though they are non-clinical and not meaningful to all end-users (See Fig. 2). Such endpoints do not directly reflect or describe how a patient feels, functions and survive and therefore hold no intrinsic value to patients, but are nonetheless important because they are strongly associated with a meaningful outcome, and therefore compellingly influence clinical decision making, for example a troponin result or a measured blood pressure.

Trial endpoints may be used to derive metrics which are used to further evaluate the impact of an intervention, for example from a population or policy perspective, such as number needed to treat or harm, or the incremental cost per quality-adjusted-life-years gained. These metrics are important from a societal, and consequently translation perspective [3].

3.2. Surrogate endpoints

Surrogates are those endpoints that do not directly measure how a person feels, functions or survives, but which are so closely associated
with a clinically meaningful endpoint that they are taken to be a reliable substitute for them [2]. The quality of a surrogate endpoint is therefore determined by the extent to which a treatment effect on that surrogate corresponds to a treatment effect against one or more clinically meaningful outcomes. Conceptually, the best surrogate endpoints directly measure causal intermediaries of the effect of an intervention on a clinically meaningful outcome, where essentially all effects on that outcome are mediated through that intermediary, and where there is little attenuation between the effect of a treatment on the intermediary and the intermediary’s effect on the clinically meaningful outcome [5]. Surrogates which do not causally influence the meaningful outcome may still be statistically associated with it for a given treatment in a given context, but this association may not generalise well to other clinical contexts, populations or interventions.

A validated surrogate is one which reliably captures a treatment effect against one or more clinically meaningful endpoints, bearing in mind that the strength of this association may be context dependent, and reliability cannot be inferred unless there are multiple randomised, controlled trials of interventions that have the same or similar effect on both the surrogate and the clinically meaningful end-point [5]. The US Food and Drug Administration (FDA) provides a list of validated and unvalidated surrogates [6], for example HbA1c is listed as a marker of risk of long term microvascular complications in type 2 diabetes mellitus. However, unvalidated surrogates are sometimes selected for lack of a validated surrogate, and there is no standardised process or agreed criteria that must be met for validation.

Prentice first described the criteria for scientific validation [7], proposing the surrogate should be statistically correlated with the clinical outcome of interest, and also fully capture the effect of the intervention on the outcome. The latter criterion has been critiqued as being too stringent [5]. Fulfilment of the Prentice criteria requires an understanding of the causal pathways of disease and the effects of an intervention on this pathway, and such complexities might never be confidently understood entirely. Surrogates typically only capture ‘on-target’ effects, that is effects that are anticipated based on our understanding of the causal pathway of the disease process; ‘off-target’ effects of an intervention lie outside this causal pathway, and are therefore unanticipated, and may not be captured by a surrogate, but may nonetheless impact importantly (positively or negatively) on the meaningful outcome [5]. Alternative approaches for validation of surrogates have been described elsewhere [8,9].

Fleming & DeMets warn that even if surrogates correlate with an outcome of interest, they may fail to predict clinical endpoints through one of three mechanisms [10]. The first is failure of the surrogate to lie on the causal disease pathway. An example is the use of laboratory measures to evaluate the impact of HIV treatment in pregnancy to reduce mother to child transmission of HIV infection [10]. The maternal CD4 count and HIV viral load are both statistically correlated with the risk of transmission in untreated women; low CD4 count and high viral load are both associated with increased risk. HIV viral load, which measures the amount of circulating virus in the mother’s blood, is thought to lie on the causal pathway between treatment and transmission because circulating virus is thought to be a prerequisite for transmission. Any treatment that reduces the maternal viral load can therefore reasonably be expected to reduce the risk of transmission. The CD4 count however, which measures the status of the mother’s immune system, may not be causally related to transmission. Instead, high viral load in untreated women causes low CD4 count, so the association between low CD4 count and risk of transmission may be confounded by the higher viral loads in women with low CD4 count. This means that treatments that impact on CD4 count (and not the viral load) may not influence the risk of transmission. HIV viral load is therefore prima facie a more reasonable surrogate than the CD4 count for capturing the effect of maternal interventions on risk of mother to child transmission.

The second reason for failure of a surrogate is the existence of more than one causal pathway impacting on the outcome, where the surrogate lies on one pathway only [11]. In the above example, maternal viral load might only be a reasonable surrogate for mother-to-child-transmission for those treatments that mediate their protective effects by inhibiting viral replication. Caesarean section is also protective against transmission, but through alternative pathways, presumably by decreasing exposure of the newborn to maternal blood and secretions. Maternal viral load would not be expected to be a useful surrogate in that context.

Finally, the intervention may produce off-target effects that impact on the measured outcome [11]. The Cardiac Arrhythmia Suppression Trial (CAST) was designed to test the hypothesis that suppression of...
asymptomatic or mildly symptomatic ventricular arrhythmias with anti-arrhythmic agents (flecainide or encaïnide) would reduce the risk of death or cardiac arrest requiring resuscitation in survivors of myocardial infarction [12]. Although the pilot study for this trial found these agents suppressed arrhythmias adequately in the target population [13], mortality increased 3-fold in the CAST owing to effects of these drugs on mortality through alternative pathways, possibly through unanticipated pro-arrhythmic effects [12], prompting withdrawal of these drugs from the market [5].

3.3. Endpoint characteristics: what is ideal?

Conceptually, an ideal endpoint should be a valid and applicable measure of how a patient feels, functions or survives [2] and be perceived by end-users of the research as having meaning and value. To be valid, an endpoint should capture the outcome of interest accurately (measure what is intended), precisely (with minimal error or uncertainty) and consistently with repeated measurements [14]. This is easiest to achieve when the outcome of interest can be measured directly, such as death. An ideal endpoint should also be measured easily, without additional risk, at low cost, at minimal inconvenience to the patient [15], and, if possible, captured as part of routine data collected as part of clinical care. Death is one example of an endpoint for interventions of highly fatal conditions which fulfils all these criteria, including the fact that this endpoint is meaningful to all end-user groups. For the majority of conditions where death is rare, or where survival may be associated with significant suffering or disability, death will not capture all relevant and meaningful outcomes.

Standardization of endpoints is increasing through the development and adoption of core outcome sets [16]. Core outcomes are the effect(s) of a health intervention which are agreed as being important to end-users, including patients. A core outcome set (COS) is a minimum agreed list of outcomes that should be measured and reported in trials [17]. COS are disease-, population- and/or intervention-specific, however there is often considerable overlap between outcomes selected across different research domains given that outcomes are likely to be important irrespective of the underlying disease process. Guidelines are available to inform development of core outcomes sets and identification of optimal methods for outcome measurement [16,17]. Patients, clinicians, policy-makers, industry representatives, and members of the public may be involved in the development of core outcome sets depending on existing subject matter knowledge, the rationale for development, and feasibility constraints [16–18].

3.4. Evolution of endpoints to capture different treatment effects

It may be helpful to consider when and why the use of different endpoint types has evolved over time; this is summarized in Fig. 2 [19–22]. Because interventions impact patients in different ways and may have more than one consequence (positive or negative), decision making around the use of an intervention should consider the net benefit versus risk [17]. Increasingly complex endpoints have evolved in parallel to advances in trial design and data capture in order to assess multiple important effects of an intervention in aggregate, or to determine whether the intervention is likely to have a net benefit to a patient overall. In the current era of patient-centered healthcare, individualised endpoints have also been recently proposed as a framework for evaluating personally defined risk and benefit [23].

No endpoint type is universally better than all others, but rather, the different characteristics and properties of each type make them better suited for use in different contexts; this is considered in further detail below. A summary of the strengths and limitations of various types of endpoints described in the literature are presented in Table 1 [14,19,21,24].

3.5. Multiple and combination primary endpoints

Multiple or combination primary endpoints may be required to capture the aggregate risk-benefit effect of an intervention [3,25]. This may be considered when multiple disparate outcomes have comparable importance, if each of those outcomes are individually rare, or if no consensus can be reached regarding which is most important [3].

3.6. Multiple endpoints

Multiple endpoints can be chosen and evaluated separately, such that a significant treatment effect against any one of the endpoints may be taken as evidence of efficacy. This approach may be useful in diseases that have multiple sequelae, where improvement in any pre-specified endpoint is clinically meaningful even in the absence of improvement in any other [3,19]. Because the risk of type I error increases with every additional endpoint assessed, appropriate statistical adjustments for multiplicity are generally needed to contain the risk of a false positive trial result; regulatory authorities are particularly focussed on this issue and have given guidance on managing this risk [3].

Multiple primary endpoints become ‘co-primary’ if an effect on multiple outcomes is required to demonstrate proof of efficacy [3]. An example of co-primary endpoints includes both cognitive and functional assessments in studies of Alzheimer’s disease [3,26] in which for a treatment to be considered efficacious it must demonstrate a beneficial effect on both cognition and function. There is no risk from multiplicity when co-primary endpoints are used [3]; conversely, the power of a study is typically diminished by the requirement to demonstrate significant efficacy against more than one endpoint, unless those endpoints are highly correlated.

3.6.1. Combination endpoints

Combination endpoints may be either composite or multi-component [3,19].

3.6.1.1. Composite endpoints. Some trials combine measures of multiple outcomes (such as death and major morbidity events) into a single measure of effect, or composite endpoint [3]. This helps to avoid the multiplicity issues inherent when multiple endpoints are assessed separately [14,21]. Composite endpoints are sometimes used to aggregate the total benefit when the goal of therapy is to prevent or delay a number of important but uncommon clinical events [21]. One example is a composite endpoint which comprises any of death, myocardial infarction, stroke or revascularisation in cardiovascular trials [3].

The value of composites is influenced by the relative importance of its components. The components of a standard composite endpoint are implicitly ascribed equal weight. If the components do not have comparable importance (for instance death and revascularisation) [27] the trial results may be difficult to interpret and less useful for end-users unless the size or direction of the treatment effect against each component is uniform. Individual components of the composite must be reported separately (as secondary endpoints) in addition to the overall result, but there may be insufficient power to determine the treatment effect for each component. An additional limitation of composite endpoints is that repeated (and possibly more serious) events are ignored [28].

There are two broad approaches to the analysis of composite endpoints. The ‘first combine and then compare’ method involves combining the components into a single composite endpoint and then comparing the frequency or rate of the composite between treatment and placebo groups [23,29]. The second method, to ‘first compare and then combine’ or the ‘win-ratio’ approach, is gaining traction as an alternative method which helps to account for heterogeneity in treatment effects across the component outcomes [30]. This involves matching pairs of patients in the treatment and placebo arms based on...
their risk of experiencing the outcome of greatest importance included in the composite (such as death), and examining component outcomes in a prioritised fashion. This creates an implicit weight between the component outcomes of the composite, but doesn’t consider their exact weighting; if assessment of the first outcome included in the composite results in a tie (or doesn’t occur in either group), the second most important outcome is evaluated. The number of ‘wins’ versus ‘losses’ is then compared between groups to calculate the win ratio [31]. Pocock et al. [31] have applied this method to the CHARm trial results, which evaluated use of an ACE inhibitor compared to placebo in chronic heart failure using a composite endpoint incorporating death, myocardial infarction, cerebrovascular accident (CVA) and target vessel revascularisation showed CABG to be superior to PCI in 1204 propensity-matched patients at 3 years. Weighting the component outcomes according to clinical significance, however (with death considered worse than CVA, followed by MI and finally revascularisation) found no significant difference between CABG to be superior to PCI in 1204 propensity-matched patients at 3 years.

\[ \text{Win ratio} = \frac{\text{Number of wins}}{\text{Number of losses}} \]

**3.6.2. Multi-component endpoints**

A multi-component endpoint combines numerous pre-specified component outcomes into a single score or rating which is calculated using a multi-attribute instrument, where the scores for each attribute may be either weighted or unweighted \([3,32]\). In contrast to composite endpoints, the components in a multi-component endpoint may not be meaningful when analysed individually, and all components must be assessed for each participant and contribute to the overall score. Unweighted multi-attribute instruments effectively assign equal importance to all items, and an overall score is obtained by simply summing the responses, such as psychometric assessments that measure cognitive ability.

**3.7. Weighting and utility**

The individual components included in composite and multi-component endpoints often don’t have comparable importance. Weighted analysis has been proposed as one method for overcoming this issue \([33]\), where the weights are intended to reflect the relative importance of an individual outcome relative to others \([20]\).

Weights may be assigned by expert judgement (obtained through a Delphi panel process, for example) \([32]\) or elicited using either ‘stated’ or ‘revealed’ preference methods \([34]\). Stated preferences are derived from decisions made by individuals when confronted with realistic, hypothetical choice scenarios. Time-trade-off, standard gamble, visual analogue scales (where a specific health state is rated on a scale from 0 to 100), and discrete choice experiments (DCEs) are examples of stated preference techniques \([34,35]\). Revealed preference methods assign weights to outcomes based on observed choices made by individuals in real-life scenarios. Most obtain individual participant preference information, although disability-adjusted life years (DALY) is one method which reflects weighting of health outcomes obtained at a population level \([36]\).

Weighted composite endpoints have been used extensively in cardiovascular research. One example is a post hoc analysis of the DELTA trial \([37]\). This study evaluated the impact of either percutaneous coronary interventions (PCI) or coronary artery bypass grafting (CABG) in patients with left main coronary artery disease. Using a primary composite endpoint incorporating death, myocardial infarction, cerebrovascular accident (CVA) and target vessel revascularisation showed CABG to be superior to PCI in 1204 propensity-matched patients at 3 years. Weighting the component outcomes according to clinical significance, however (with death considered worse than CVA, followed by MI and finally revascularisation) found no significant difference between revascularisation strategies.

Utilities are sometimes applied to individual outcomes (including health state descriptions) in which the health state outcome is converted to a utility measure. Utilities attempt to quantify the desirability or value of an outcome or health state, and specifically how much better/worse one is over another \([32]\). Assigning a utility value to a range of possible health outcomes enables calculation of a single overall utility score for each participant which can then be aggregated over all participants in each study arm.

---

**Table 1**

Strengths and limitations of trial endpoints.

<table>
<thead>
<tr>
<th>Endpoint Type</th>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Singular</strong></td>
<td>Routinely collected information</td>
<td>Doesn’t consider that an intervention may impact on the patient in different ways</td>
</tr>
<tr>
<td><strong>Clinically observed</strong></td>
<td>Typically well-accepted approach by scientific community</td>
<td>May need supportive secondary analyses to be persuasive</td>
</tr>
<tr>
<td><strong>Surrogates</strong></td>
<td>Reduction in sample size</td>
<td>May fail to predict clinically meaningful endpoints</td>
</tr>
<tr>
<td></td>
<td>Shorter trial duration</td>
<td>May not be sensitive to change at all stages of disease</td>
</tr>
<tr>
<td></td>
<td>Decreased cost of trial</td>
<td>Validation process often challenging</td>
</tr>
<tr>
<td></td>
<td>Accelerated approval/dissemination of effective therapies</td>
<td>Therapeutic advances may alter the validity of the surrogate measure</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Endpoint Type</th>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Multiple or combination</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Co-primary</strong></td>
<td>Useful if more than one important outcome exists &amp; demonstration of 1 is enough to support clinical efficacy</td>
<td>Adjustment for Type 1 error is required.</td>
</tr>
<tr>
<td><strong>Co-primary</strong></td>
<td>Useful if demonstration of two or more outcomes is necessary to establish clinical benefit</td>
<td>Hard to interpret if results occur in different directions</td>
</tr>
<tr>
<td><strong>Composites</strong></td>
<td>Improves statistical efficiency and precision</td>
<td>Implementation may be complex and resource-intensive.</td>
</tr>
<tr>
<td></td>
<td>Increases power (reduces sample size requirement)</td>
<td>Components may be inappropriately combined or reported.</td>
</tr>
<tr>
<td></td>
<td>Ability to measure small effects</td>
<td>May be difficult to interpret study findings and determine which of the component endpoints are impacted by the intervention; the effect is often smallest for the most important component and biggest for the less important components.</td>
</tr>
<tr>
<td></td>
<td>Lower cost</td>
<td>Prone to post-hoc analyses/bias.</td>
</tr>
<tr>
<td></td>
<td>Earlier trial completion</td>
<td>Key data often missing or unclear.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Endpoint Type</th>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Multi-component endpoints</strong></td>
<td>Allows single evaluation of numerous components without creating multiplicity issues</td>
<td>Adjustment for Type 2 error is required.</td>
</tr>
<tr>
<td><strong>Weighted endpoints</strong></td>
<td>More complex/robust evaluation of the effectiveness of treatment intervention(s) that considers the relative importance of components</td>
<td>Can lose meaning if components of composite move in opposite directions</td>
</tr>
<tr>
<td><strong>Endpoints that are participant specific</strong></td>
<td>Best reflects clinical decision-making</td>
<td>Individual components may not have clear meaning.</td>
</tr>
<tr>
<td></td>
<td>Theoretically would represent the gold-standard for informing personalised, evidence-based medicine.</td>
<td>If components aren’t concordant, study power may be compromised</td>
</tr>
<tr>
<td></td>
<td>May result in increased power to detect real treatment effects for patients</td>
<td>Process of assigning weights not standardised, and can involve lengthy processes.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>May be costly</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Endpoint Type</th>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Utility</strong></td>
<td>Complex; logistically difficult.</td>
<td>Generalisability of trial results may be limited.</td>
</tr>
<tr>
<td></td>
<td>Requires large data capture</td>
<td>Requires large data capture</td>
</tr>
</tbody>
</table>
Discrete choice experiments (DCEs) are one method for calculating utility, and are relatively new to health, but growing over time [38]. DCEs present participants with a series of hypothetical scenarios, called choice sets, asking them to make decisions about preferred health outcome states. This process requires respondents to make trade-offs between different aspects health-related status (like benefit versus drug toxicity) which enables the quantification of the relative importance of outcomes, which is superior to other techniques which simply rank or rate them [39]. Quality of life tools, which capture an individual’s subjective assessment of their physical, mental and social wellbeing are one example of utility instruments [34,35,39]; they may be generic, such as the EuroQol five-dimensional questionnaire (EQ-5D) and the six-dimensional health state short form (SF-6D) [40] or disease-specific, such as the Cystic Fibrosis Questionnaire-revised [41].

3.8. Participant specific endpoints

Significant heterogeneity exists in individual preferences for outcome states, even among those with the same disease and similar baseline health states [3]. End-users may want to individualise treatment decisions according to personal characteristics including disease stage and comorbidities, the availability or lack of treatment options, values and beliefs, and financial considerations [14]. Consequently, it may be difficult to directly apply trial evidence to inform patient management if the patient at hand differs demographically or clinically from average trial participants, or if the endpoints selected for the trial are of secondary or of no importance to the patient. For example, a trial which reports the efficacy of a drug for return to work, may have little applicability for a retired patient who desires a return to independent living.

Individualised endpoints for participants in a trial may provide a framework for evaluating personally-defined risk and benefit. Two approaches have been proposed. The first [42] employs a sliding dichotomy which defines treatment success for a given trial participant based on what experts deem to be achievable and desirable given their baseline disease status and prognosis. This allows patients to be enrolled into a trial across a spectrum of baseline health and disease severity (such as stroke severity), with all participants able to contribute to the analysis.

Iwashyna et al. [23] has built on this prognostic stratification by proposing use of a ‘values clarification instrument’ to elicit preferences from prospective participants in a trial for a range of possible health outcome states. These preferences might be combined algorithmically prior to randomisation to determine which endpoints are both achievable and most desirable to patients and treatment success could then be defined as the realization of one or more of these endpoints. This approach might be difficult to apply in trials in acute care settings, but might be applicable in non-acute settings, like trials of interventions in chronic diseases like cystic fibrosis.

4. Discussion

The range and complexity of endpoint types available for use over time has increased alongside the evolution of increasingly complex trial designs. Increasing recognition of the need to engage different end-user groups in endpoint determination is an important step towards improving the value of research that is conducted and the likelihood of translation of research findings. Whilst significant developments have occurred in this field, clearly the optimisation and selection of endpoints for clinical trials remains a science in evolution.

Whilst the CONSORT statement provides guidance about how to report outcomes for randomised controlled trials [43], no guidance is available to inform optimal selection of endpoints, nor methodology available to quantify endpoints as best, and so endpoint selection will likely remain the prerogative of those who design and sponsor trials. The exception to this is for trials to support licensure of new therapeutic products, where the regulator may be directly involved in nominating the endpoints. Such trials should provide evidence to guide practice, however registration endpoints do not always capture outcomes that are clinically meaningful to patients or clinicians, and consequently may fail to achieve this goal. There is increasing recognition of the need for regulatory authorities to consider patient preference information when stipulating endpoints [3].

Outside the context of licensure, an end-point can be regarded as having some value if there is a history of end-users changing practice or policy for trials that have that end-point. Where there is no history there may be value in pre-trial surveys or focus groups of end-users to establish that proposed end-points would be regarded as sufficient to change practice or policy.

Failure to involve all end-user groups in discussion about endpoint suitability can compromise the translation of the results of a trial into clinical practice. We believe this is common and an important oversight. Assumptions made during the design of clinical trials can contribute to this situation. Firstly, clinicians often make incorrect assumptions about patient preferences and what patients value [44]. Secondly, the broader community of clinicians and policy-makers are often not engaged to establish that a selected end-point meets their requirements. While it can be difficult to select endpoints that will satisfy all end-users, at least understanding the perspectives and priorities of these groups will help avoid the use of end-points that are misaligned with the trial’s objectives. In trials examining optimal treatment of pulmonary exacerbations of cystic fibrosis for example, clinicians might most value the impact of treatment on measurements of lung function (change in FEV\textsubscript{1.0} from baseline) [unpub, Snelling], patients may attach greater significance to the effect of treatment on functional status or quality of life, while policymakers might prioritise the cost-effectiveness of external strategies like antimicrobial resistance implications of specific treatment regimes. An empiric understanding of alternative endpoint types can assist in the design of trials to meet the requirements of all end-users. Achieving consensus between disparate groups of end-users about meaningful outcomes is more likely to occur when research questions arise in the context of goal-orientated care, where treatment is administered based on what outcomes are considered achievable and desirable to patients [44].

Endpoints may also be limited to the extent to which they are context- (including timing), patient- and intervention (including drug class)- specific [14]. Endpoints may not be applicable in settings that do not have the capability of performing the selected outcome measurement(s). Endpoints might not reliably measure outcomes in all individuals included in the study population, which produces ceiling and floor effects that must be considered [45]. For example, spirometry is a valuable measure of lung function in adults but cannot be reliably performed in children less than 6 years old or in patients with end-stage respiratory failure. Endpoints may also not consistently detect the effect of a treatment intervention across all stages of disease. Forced expiratory volume in 1 s (FEV\textsubscript{1.0}), a marker of lung function, and radiological studies are both insensitive markers to change in lung disease early in cystic fibrosis, for example [14]. Conversely the 6-min walk test is not applicable for patients with muscular dystrophy who are already confined to a wheelchair. Drugs may impact on the same outcome via different pathways; for example, anti-arrhythmic and lipid-lowering agents impact on cardiovascular mortality through different mechanisms. This means use of surrogates across different classes of drugs, even when used for a similar purpose, cannot be assumed to be appropriate [5].

Where surrogates are used as endpoints, it may be unclear what degree or duration of effect corresponds with a clinically meaningful effect [5]. While it is widely agreed that lowering blood pressure is causally associated with reduction in the risk of cardiovascular death, it may not be easy to translate exactly how a given reduction in blood pressure over a given period of time translates into a quantifiable reduction in mortality.

Selecting endpoints which meet the conceptual ideal is challenging.
and may not be possible, forcing researchers to compromise and make pragmatic decisions about those selected. Further, health outcomes identified as meaningful to end-users may be abstract, and therefore may not be directly measurable. In this regard, the availability (or absence) of appropriate scales of measurement for outcome evaluation may be an important factor that drives endpoint selection [23].

5. Conclusions

Optimisation and selection of endpoints for clinical trials is an evolving field. Given the purpose of late phase trials is to inform clinical practice and policy, endpoints should measure outcome(s) which are meaningful to end-users that reflect or describe how a patient feels, functions or survives. Understanding the range of endpoints available for use and the context in which they have arisen together with their strengths and limitations will help inform end-users when selecting endpoints for late phase trials.

Future work should focus on streamlining processes for identifying prioritised outcomes for different end-user groups across different research domains and on developing a methodology for qualifying endpoints as best. There is a need for universally agreed guidelines to inform optimal selection and reporting of endpoints; such guidelines should emphasise the importance of endpoints being suited to the trial purpose and participants and acceptable to relevant end-user group(s).

Justification of the selection of endpoints in all trials should be reported. Such guidelines may be beneficial for end-users and help reduce research waste.

Funding

This work was supported by a Perth Children’s Hospital Foundation Grant (#9757). CM is supported by fellowship grants from the National Health and Medical Research Council (NHMRC; GNT1150996), the Wesfarmers Centre, and the Perth Children’s Hospital Foundation (#9772). TS is supported by a Career Development Fellowship from the NHMRC (GNT1150996). Sponsors had no role in the design, collection or synthesis of data, the writing of this review, or the decision to submit for publication.

Declaration of competing interest

The author(s) declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgements

The authors wish to thank senior librarians Karen Jones and Samantha Blake located at the J. Robin Warren Library (UWA, Perth) for their guidance which informed the search strategy employed in this review.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.conctc.2019.100486.

References


Chapter 4: Outcomes and endpoints reported in studies of pulmonary exacerbations in people with CF

4.1 Chapter summary

This chapter systematically reviews the outcomes and their corresponding endpoints that have been reported in studies of pulmonary exacerbations of CF, and distinguishes those that are likely to be meaningful based on those that capture, *prima facie*, how a person feels, functions or survives.

The outcomes identified in this review, together with additional outcomes identified by people affected by CF, were subsequently prioritised and considered for inclusion in a discrete choice experiment (DCE) (*Chapter 6*). The results of the DCE are presented in *Chapter 7*.

Death, quality of life, patient-reported outcomes, the forced expiratory volume in one-second \([\text{FEV}_1]\) and structural lung changes were identified as outcomes that are most likely to be meaningful. Development of a core outcome set (COS) in collaboration with key stakeholders is recommended (*Chapter 9*).

4.2 Journal article


Appendix 1: Abbreviations and their meanings

- **S1**: Search strategy (Medline)
- **S2**: Search strategy (Embase)
- **S3**: Published studies that met inclusion criteria
- **S4**: Strengths and limitations of identified outcomes

Appendix and supplementary materials are available [here](http://example.com).
Review

Outcomes and endpoints reported in studies of pulmonary exacerbations in people with cystic fibrosis: A systematic review

Charlie McLeod a,⁎, Jamie Wood d,⁎, André Schultz f,g, Richard Norman h, Sherie Smith i, Christopher C. Blyth a,c,l, Steve Webb k,l, Alan R. Smyth j, Thomas L. Snelling m,n

a Wesfarmers Centre for Vaccines and Infectious Diseases, Telethon Kids Institute, 15 Hospital Ave, Nedlands WA 6009, Australia
b Division of Paediatrics, Faculty of Medicine, University of Western Australia, 35 Stirling Hwy, Nedlands 6009, Australia
^c Infectious Diseases Department, Perth Children’s Hospital, 15 Hospital Ave, Nedlands 6009, Australia
^d Physiotherapy Department, Sir Charles Gardiner Hospital, Hospital Ave, Nedlands 6009, Australia
^e Abilities Research Center, Department of Rehabilitation and Human Performance, icahn School of Medicine at Mount Sinai, New York, United States of America
f Centre for Respiratory Health, Telethon Kids Institute, University of Western Australia, 35 Stirling Hwy, Nedlands 6009, Australia
g Respiratory Department, Perth Children’s Hospital, 15 Hospital Ave, Nedlands 6009, Australia
h School of Public Health, 400 Curtin University, Kent St, Bentley 6102, Australia
i Evidence Based Child Health Group, University of Nottingham, Queen’s Medical Centre, Nottingham NG7 2UH, United Kingdom
j PathWest Laboratory Medicine WA, QEII Medical Centre, Nedlands 6009, Australia
k St John of God Hospital, 12 Salado Road, Subiaco 6008, Australia
l School of Population Health and Preventive Medicine, 553 St Kilda Rd, Monash University, Melbourne 3004, Australia
m, n Menzies School of Health Research, PO Box 41096 Casuarina NT 0811, Australia

A R T I C L E   I N F O

Article history:
Received 27 May 2020
Revised 25 August 2020
Accepted 25 August 2020
Available online 13 November 2020

Keywords:
Cystic fibrosis
Pulmonary exacerbation
Outcome assessment
Patient reported outcome measures
Outcome variables
Endpoint measure

A B S T R A C T

Background: There is no consensus about which outcomes should be evaluated in studies of pulmonary exacerbations in people with cystic fibrosis (CF). Outcomes used for evaluation should be meaningful; that is, they should capture how people feel, function or survive and be acknowledged as important to people with CF, or should be reliable surrogates of those outcomes. We aimed to summarise the outcomes and corresponding endpoints which have been reported in studies of pulmonary exacerbations, and to identify those which are most likely to be meaningful.

Methods: A PROSPERO registered systematic review (CRD42020151785) was conducted in Medline, Embase and Cochrane from inception until July 2020. Registered trials were also included.

Results: 144 studies met the inclusion criteria. A wide range of outcomes and corresponding endpoints were reported. Death, QoL and many patient-reported outcomes are likely to be meaningful as they directly capture how people feel, function or survive. Forced expiratory volume in 1-second (FEV1) is a validated surrogate of risk of death and reduced QoL. The extent of structural lung disease has also been correlated with lung function, pulmonary exacerbations and risk of death. Since no evidence of a correlation between airway microbiology or biomarkers with clinically meaningful outcomes was found, the value of these as surrogates was unclear.

Conclusions: Death, QoL, patient-reported outcomes, FEV1, and structural lung changes were identified as outcomes that are most likely to be meaningful. Development of a core outcome set in collaboration with stakeholders including people with CF is recommended.

© 2020 Published by Elsevier B.V. on behalf of European Cystic Fibrosis Society.

⁎ Corresponding author: Charlie McLeod; Infectious Diseases Department, Perth Children’s Hospital, 15 Hospital Avenue Nedlands, WA 6009.
E-mail addresses: charlie.mcleod@health.wa.gov.au (C. McLeod), Jamie.Wood@health.wa.gov.au (J. Wood), Andre.Schultz@health.wa.gov.au (A. Schultz), richard.norman@curtin.edu.au (R. Norman), Sherie.Smith@nottingham.ac.uk (S. Smith), Christopher.blyth@uwa.edu.au (C.C. Blyth), steve@stevewebb.com.au (S. Webb), Alan.Smyth@nottingham.ac.uk (A.R. Smyth), tom.snelling@sydney.edu.au (T.L. Snelling).
https://doi.org/10.1016/j.jcf.2020.08.015
1569-1993/© 2020 Published by Elsevier B.V. on behalf of European Cystic Fibrosis Society.
1. Introduction

There is no consensus core outcome set (COS) of outcomes and corresponding endpoints which should be used in studies evaluating pulmonary exacerbations in people with cystic fibrosis (CF). Pulmonary exacerbations drive lung damage and are characterised by acute worsening of pulmonary status [1]. Treatment for these episodes generally involves a combination of antimicrobial, anti-inflammatory and airway clearance therapies (including physiotherapy and mucolytics) and optimisation of nutrition [2]. Despite over 667 trials and 43 Cochrane reviews of treatment of pulmonary exacerbations, there is no agreement regarding optimal management [3]. In part, this may be attributed to the inconsistent reporting of outcomes and endpoints in these studies and the selection of outcomes that may not be meaningful to people living with disease. To be meaningful, outcomes and their corresponding endpoints should capture how people feel, function and survive and be acknowledged as important to people living with disease [4], or should be reliable surrogates of those outcomes.

Core outcome sets are collections of agreed outcomes derived by broad stakeholder consensus that should be measured and reported in all studies for a specific condition [5–7]. The Core Outcome Measures in Effectiveness Trials (COMET) initiative was established in 2010 to improve consistency in the selection of meaningful outcomes when designing studies, to avoid duplication of research, facilitate collaboration, and improve the value of research that is conducted [8].

The primary aim of this review was to systematically identify the range of outcomes and corresponding endpoints that have been reported in studies involving treatment of pulmonary exacerbations in people with CF, and to identify those which directly capture how people feel, function and/or survive, or which may be reliable surrogates of those outcomes. We anticipated that many of the outcomes and endpoints that have been reported would not meet our criteria for being considered meaningful. Additional objectives were to summarise the reported strengths and limitations of these outcomes and endpoints and to describe how their use has changed over time. This review is targeted toward end-users involved in the design and conduct of studies in people with CF; here we identify a subset of outcomes and endpoints that are most likely to be meaningful and which should therefore be considered for inclusion in a COS. Further steps towards developing a COS for studies of pulmonary exacerbations will require collaboration with relevant stakeholders including people with CF. It is hoped that a COS will improve the value of research that is conducted in this field and contribute to better outcomes for people living with disease.

2. Methods

2.1. Search strategy and selection criteria

The search strategy is provided in Tables S1 & S2. We searched MEDLINE, Embase and the Cochrane databases from inception until July 2019. Trials identified from the Clinical Trials and the European Clinical Trials registries were also included.

Inclusion criteria were studies written in English evaluating outcomes in pulmonary exacerbation studies in people with CF of all ages, including observational studies and clinical trials, reviews and abstracts. Registered, unpublished trials proposing novel outcomes and endpoints were also evaluated. References in selected articles that provided additional information of the correlation between surrogate endpoints and clinically meaningful outcomes were also reviewed. Given this work was designed to inform a COS for late phase efficacy trials of interventions for CF pulmonary exacerbations, phase I and II trials and pharmacokinetic studies were excluded.

Outcomes were defined as those patient characteristics or biological processes that are targeted for improvement by a trial intervention in individual participants (e.g. lung function), and endpoints as the specific analysed parameter(s) corresponding to those outcomes (e.g. change in the percentage predicted forced expiratory volume in one-second [ppFEV₁] from baseline to day 10) [9]. Composite endpoints were specified. Outcomes and endpoints were categorised as clinical or non-clinical. Clinical outcomes were categorised as (i) clinician reported outcomes where they involved judgement or interpretation of clinical signs or events (such as a pulmonary exacerbation event) (ii) standardised performance measures (e.g. 6-minute walk test) (iii) patient reported outcome(s), or (iv) observer-reported outcome(s) (e.g. weight or height). Non-clinical endpoints (including biomarkers) were defined as measures of an underlying biological or pathologic process [10].

The search strategy was independently executed by two authors (CM & JW). Potentially eligible studies were downloaded to Endnote by CM, and duplicates removed. Full text articles were retrieved and eligibility confirmed by both reviewers. If full text manuscripts were unobtainable but relevant data were available in the abstract, these were included. Otherwise, articles were excluded. A third reviewer (JS) determined eligibility where there wasn’t consensus. Relevant data were extracted by CM and recorded in an Excel database and confirmed by JW.

The number of published trials reporting each outcome were recorded. While reviews and systematic reviews were included in this study to capture the range of outcomes and endpoints reported, these provided inconsistent data about the outcomes studied in individual trials and were consequently excluded from this calculation. An assessment of the quality of included studies including a risk of bias assessment and meta-analysis of data was not performed as it was deemed a priori that these steps would not be required to meet the objectives of this review and would not alter the study findings.

2.2. Abbreviations and meanings

Appendix 1 provides a full list of abbreviations and their meanings used throughout this manuscript and the supplementary materials.
3. Results

3.1. Outcomes and endpoints

The search strategy is depicted in Fig. 1. A summary of published studies that met inclusion criteria is included in Table S3.

Of the published studies, 127 met inclusion criteria; four of these were identified from trial registries; of these, one duplicate article was found. Sixteen registered studies proposing novel outcomes and endpoints for evaluation without published results were also included. One hundred and eleven full-text manuscripts were obtained. Of the remaining 17 articles, all abstracts except one [11] contained relevant data and were included.

Tables 1 and 2 [4] summarise the clinical and non-clinical outcomes and corresponding endpoints that were identified as meaningful. A more detailed summary including their reported strengths and limitations is found in Table S4.

The following clinical outcomes were identified: death, a composite of pulmonary exacerbation or death, mechanism or biological outcomes, QoL, clinical scores, individual patient-reported outcomes, signs, success of therapy, clinical events including pulmonary exacerbations or adverse events, functional exercise capacity, hospitalisation and outcomes related to antibiotic therapy. Non-clinical outcomes included laboratory or radiological tests or costs of treatment.

Airflow obstruction was the most commonly reported outcome; this was predominantly measured as FEV₁, standardised for age, sex and height, and analysed as a change in the absolute or percentage predicted value between two points in time [3,5,12–86]. The baseline FEV₁ was either the FEV₁ at initiation of intensive therapy or the ‘best’ value within the preceding 3–12 months; this was variously compared to the FEV₁ at 7, 10–14 days or 1–3 months after treatment cessation [12,29,31,42,71,75,87–91]. A pre-defined minimum clinically important difference (e.g. 10% improvement in FEV₁) was sometimes used to define treatment success [29,42,71]. The number of pulmonary exacerbations or the time to the next exacerbation during a defined period were alternative endpoints reported, including in studies of young children incapable of performing spirometry [90]

A change in symptoms and/or signs and/or radiological changes captured as a single clinical score between two points in time was another common endpoint used to evaluate treatment effects. These scores were based on patient-reported outcomes and/or input from clinicians (see Table S4). The CF Respiratory Symptom Diary /Chronic Respiratory Infection Symptom Score (CFRSD-CRISS) was the most commonly used; this comprised the first eight items of the CFRSD questionnaire including breathing difficulty, cough, sputum production, chest tightness, wheeze, fever, tiredness and presence of absence of chills or sweats and converted the result to a score between 0 and 100 [2]. The longer CFRSD (16-item questionnaire) [92] has also been used to quantify symptoms and their impacts; this questionnaire was validated for use in people ≥12 years old and asked participants to recall respiratory and constitutional symptoms (including sleep) over the preceding seven days and the impact of disease on their emotional state and ability to undertake activities of daily living during this time [93]. The Schwachman-Kulczycki (SK) score (modified version SK-m 1964) was originally developed as a tool to monitor longitudinal disease progression, but has also been applied in trials of interventions for pulmonary exacerbations [57,62,63,75,78,81,92,94]; it captures four domains including general activity, physical examination, nutrition and radiological findings. Other combined clinician/patient reported scoring tools developed specifically to evaluate interventions in trials of pulmonary exacerbations were the CF Clinical Score (CFCS) [6], a scoring system proposed by Valetta [47], the Rainbow Babies and Children’s Hospital Efficacy Score [83], and the modified Huang score (1976) for use in people with end-stage CF disease [75].

The impact of treating pulmonary exacerbations has also been measured as the change in generic or disease-specific QoL scores during therapy, such as scores generated from responses to the CF Questionnaire (CFQ) [original version [1987] or revised questionnaire (CFQ-R [2000]) [95]. The CFQ-R is available in 34 languages and is validated for use in adults and children ≥ 6 years old; it aims to capture the impact of disease on overall health, daily life, perceived well-being and symptoms over the preceding two weeks, evaluating respiratory and gastrointestinal symptoms, exercise tolerance, constitutional symptoms, body image, mood,
Table 1
Clinical outcomes reported in published studies of pulmonary exacerbations in people with CF.

<table>
<thead>
<tr>
<th>CLINICAL OUTCOMES (ObsRO, ClinRO, PerfO or PRO/PROM)</th>
<th>NUMBER OF STUDIES REPORTING OUTCOME</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortality(^{\ast})</td>
<td>6 systematic reviews and 1 review excluded</td>
</tr>
<tr>
<td>Composite outcome: Death or time-to-next pulmonary exacerbation(^{\ast})</td>
<td>1, 0.9%</td>
</tr>
<tr>
<td>Mechanical or biological: ObsRO, ClinRO or PerfO(^{\ast})</td>
<td>72 (excluding 10 systematic reviews and 2 reviews), 67.3%</td>
</tr>
<tr>
<td>Airway obstruction (measures of volume e.g. FEV(_1))(^{\ast})</td>
<td>22 (excluding 4 systematic reviews), 20.6%</td>
</tr>
<tr>
<td>Airway obstruction (measure of airway flow e.g. PEFR(^{\ast}))</td>
<td>6, 3.5%</td>
</tr>
<tr>
<td>Fractional lung volumes</td>
<td>6, 3.5%</td>
</tr>
<tr>
<td>Ventilation inhomogeneity/airway obstruction</td>
<td>18 (excluding 9 systematic reviews and 1 review), 4, 3.7%</td>
</tr>
<tr>
<td>Anthropometry (including body composition)(^{\ast})</td>
<td>3, 2.8%</td>
</tr>
<tr>
<td>Energy intake &amp; expenditure</td>
<td>2, 1.9%</td>
</tr>
<tr>
<td>Physiological measures of exercise capacity</td>
<td></td>
</tr>
<tr>
<td>Respiratory physiological outcomes</td>
<td></td>
</tr>
<tr>
<td>QoL: PROM(^{\ast})</td>
<td>10 (excluding 7 systematic reviews and 2 reviews), 9.3%</td>
</tr>
<tr>
<td>QoL(^{\ast}) (Generic and disease-specific)</td>
<td></td>
</tr>
<tr>
<td>Symptoms/signs</td>
<td></td>
</tr>
<tr>
<td>Individual symptoms (sputum production, fatigue, dyspnoea, work of breathing, cough, anxiety, pain, sleep-related symptoms, urinary incontinence, activity level)(^{\ast})</td>
<td>10 (excluding 2 systematic reviews), 9.3%</td>
</tr>
<tr>
<td>Individual signs</td>
<td></td>
</tr>
<tr>
<td>Clinical scores</td>
<td></td>
</tr>
<tr>
<td>Clinical symptom +/- impact scores: PROM(^{\ast})</td>
<td>5 (excluding 1 systematic review and 1 review), 4.7%</td>
</tr>
<tr>
<td>Symptom +/- impact scores(^{\ast})</td>
<td>1, 0.9%</td>
</tr>
<tr>
<td>Composite: patient comfort, efficacy and urinary leakage(^{\ast})</td>
<td>20 (excluding 1 systematic review), 18.7%</td>
</tr>
<tr>
<td>Clinical signs/symptoms +/- radiology</td>
<td>3, 2.8%</td>
</tr>
<tr>
<td>Combined signs &amp; symptom tools(^{\ast})</td>
<td></td>
</tr>
<tr>
<td>Combined signs/symptoms &amp; radiology</td>
<td></td>
</tr>
<tr>
<td>Other PRO(^{\ast})</td>
<td>1, 0.9%</td>
</tr>
<tr>
<td>Physical activity(^{\ast})</td>
<td>1 systematic review excluded</td>
</tr>
<tr>
<td>School/work activity level(^{\ast})</td>
<td>1 systematic review excluded</td>
</tr>
<tr>
<td>Adherence &amp; patient satisfaction(^{\ast})</td>
<td></td>
</tr>
<tr>
<td>Functional measures of exercise capacity: ObsRO, ClinRO or PerfO(^{\ast})</td>
<td>5 (excluding 2 systematic reviews), 4.7%</td>
</tr>
<tr>
<td>Functional measures of exercise capacity(^{\ast})</td>
<td></td>
</tr>
<tr>
<td>Clinical events: ObsRO, ClinRO, PerfO or PRO(^{\ast})</td>
<td>14 (excluding 1 systematic review and 2 reviews), 13.1%</td>
</tr>
<tr>
<td>Pulmonary exacerbations</td>
<td>13 (excluding 3 systematic reviews), 12.1%</td>
</tr>
<tr>
<td>Treatment 'failure' vs. 'success'(^{\ast})</td>
<td>21 (excluding 3 systematic reviews), 19.6%</td>
</tr>
<tr>
<td>Treatment-related adverse-events(^{\ast})</td>
<td></td>
</tr>
<tr>
<td>Outcomes relating to provision of treatment</td>
<td></td>
</tr>
<tr>
<td>Antibiotic therapy</td>
<td>7 (excluding 5 systematic reviews), 6.5%</td>
</tr>
<tr>
<td>Hospitalisation</td>
<td>6 (excluding 7 systematic reviews), 5.6%</td>
</tr>
</tbody>
</table>

\(^{\ast}\) Directly reflect how patients feel, function or survive or reported correlation with clinically meaningful outcome.

\(^{\ast}\) ObsRO: Observer-reported outcome;

\(^{b}\) ClinRO: Clinician reported outcome;

\(^{c}\) PerfO: Performance outcome;

\(^{d}\) PRO/PROM: Patient-reported Outcome/Patient-reported Outcome Measure;

\(^{e}\) PEFR: peak expiratory flow rate;

\(^{f}\) QoL: Quality of life.

Table 2
Non-clinical outcomes reported in published studies of pulmonary exacerbations in people with CF.

<table>
<thead>
<tr>
<th>NON-CLINICAL OUTCOMES</th>
<th>NUMBER OF STUDIES REPORTING OUTCOME</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total number of studies=n/107, %</td>
</tr>
<tr>
<td>Laboratory tests</td>
<td></td>
</tr>
<tr>
<td>Systemic inflammation</td>
<td>26 (excluding 2 systematic reviews), 24.3%</td>
</tr>
<tr>
<td>Pulmonary inflammation</td>
<td>4 (excluding 2 systematic reviews), 3.7%</td>
</tr>
<tr>
<td>Immune-related</td>
<td>1, 0.9%</td>
</tr>
<tr>
<td>Arterial oxygenation</td>
<td>3, 2.8%</td>
</tr>
<tr>
<td>Treatment-related side-effects/adverse events (based on laboratory evidence)</td>
<td>21 (excluding 3 systematic reviews), 19.6%</td>
</tr>
<tr>
<td>Metabolic</td>
<td>7, 6.5%</td>
</tr>
<tr>
<td>CFTR(^{\ast}) function</td>
<td>1 systematic review excluded</td>
</tr>
<tr>
<td>Protein synthesis/turnover</td>
<td>1, 0.9%</td>
</tr>
<tr>
<td>Sputum characteristics</td>
<td>5 (excluding 1 systematic review and 1 review), 4.7%</td>
</tr>
<tr>
<td>Airway microbiology (including quantitative culture)</td>
<td>35 (excluding 9 systematic reviews and 1 review), 32.7%</td>
</tr>
<tr>
<td>Radiological</td>
<td></td>
</tr>
<tr>
<td>Structural lung damage +/- perfusion abnormalities(^{\ast})</td>
<td>12 (excluding 3 systematic reviews and 2 reviews), 11.2%</td>
</tr>
<tr>
<td>Economic</td>
<td></td>
</tr>
<tr>
<td>Cost (direct and indirect)</td>
<td>2 (excluding 3 systematic reviews), 1.9%</td>
</tr>
</tbody>
</table>

\(^{\ast}\) Directly reflect how patients feel, function or survive or reported correlation with clinically meaningful outcome.

\(^{\ast}\) CFTR: CF-transmembrane regulator.
Table 3
Outcomes that correlate with or capture how people feel, function or survive.

<table>
<thead>
<tr>
<th>OUTCOME DOMAIN</th>
<th>OUTCOME</th>
<th>AGE</th>
<th>CORRELATION WITH DISEASE-RELATED MORBIDITY/MORTALITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lung function*</td>
<td>FEV₁&lt;sup&gt;a&lt;/sup&gt; (absolute or relative change as total or percentage predicted value)</td>
<td>≥6 years</td>
<td>Low FEV₁&lt;sup&gt;a&lt;/sup&gt; strongly associated with increased mortality and decreased QoL&lt;sup&gt;b&lt;/sup&gt;. -10% improvement in FEV₁&lt;sup&gt;a&lt;/sup&gt; associated with failure to recover baseline lung function 3 months after treatment or failure to recover baseline lung function 3 months after treatment 7.8, 95% CI 1.9–31.6, p = 0.004; (OR failure to recover baseline lung function 3 months after treatment 7.8, 95% CI 1.9–31.6, p = 0.004); 95% of n = 220 recovered 90% of lost lung function in STOP trial. Structural lung disease in the first years of life is associated with inflammation and infection. Bronchiectasis is associated with absolute neutrophil count (p = 0.03), neutrophil elastase concentration (p = 0.0001) and P. aeruginosa infection (p = 0.03). CT&lt;sup&gt;c&lt;/sup&gt; scores have been shown to correlate with clinical status, lung function and disease progression and is predictive of mortality. CF-CT&lt;sup&gt;c&lt;/sup&gt; bronchiectasis (r = −0.38, p &lt; 0.001) and CF-CT&lt;sup&gt;c&lt;/sup&gt; trapped air (r = −0.35, p = 0.003) correlate with the CFQ&lt;sup&gt;c&lt;/sup&gt; respiratory symptom score. Higher modified Brody scores on CT&lt;sup&gt;c&lt;/sup&gt; imaging correlate with worse structural lung disease and scores improve after antibiotics and intensive airway clearance (p &lt; 0.001). MRI&lt;sup&gt;d&lt;/sup&gt; morphology, perfusion and global scores are sensitive to structural lung changes and pulmonary exacerbations (p &lt; 0.001) and scores normalise with antimicrobial therapy for pulmonary exacerbations (p &lt; 0.05). Significant but variable correlation with FEV₁&lt;sup&gt;a&lt;/sup&gt;/FEV&lt;sub&gt;0.5&lt;/sub&gt; in one study. One study in preschool kids showed correlation with FEV&lt;sub&gt;0.5&lt;/sub&gt;, FEV&lt;sub&gt;25–75&lt;/sub&gt; and sRAW&lt;sup&gt;e&lt;/sup&gt;. Preschool LCI&lt;sup&gt;f&lt;/sup&gt; predictor of abnormal lung function at an early school age. LCI has moderate-strong correlation with structural abnormalities on global CT&lt;sup&gt;c&lt;/sup&gt; scores. Greater ventilation inhomogeneity (higher LCI) is correlated with structural damage demonstrated on CT&lt;sup&gt;c&lt;/sup&gt; in adults/children and MRI&lt;sup&gt;d&lt;/sup&gt; (wall abnormalities, mucus plugging and abnormal perfusion p &lt; 0.05 to p &lt; 0.001). Mean LCI increased significantly during treatment for pulmonary exacerbations (by 2 units, &gt;10/1, p &lt; 0.001). Associated with loss of FEV₁&lt;sup&gt;a&lt;/sup&gt;, decreased survival and reduced QoL&lt;sup&gt;b&lt;/sup&gt;. Association between malnutrition and deteriorating lung function demonstrated. Correlates only modestly with mean absolute FEV₁&lt;sup&gt;a&lt;/sup&gt; percentage predicted change from treatment initiation (R² = 0.157, p &lt; 0.0001). The SK-m&lt;sup&gt;h&lt;/sup&gt; score correlates well with percent predicted values for FVC&lt;sup&gt;i&lt;/sup&gt; (r = 0.69) and forced expired volume in 1 s (FEV₁&lt;sup&gt;a&lt;/sup&gt;) (r = 0.67). Correlated with disease activity; significantly significant increase from stable to exacerbation state.</td>
</tr>
<tr>
<td>Structural lung disease*</td>
<td>Structural lung disease (e.g. bronchiectasis, mucus plugging, degree of air trapping, airway inflammation)</td>
<td>Birth to school age</td>
<td></td>
</tr>
<tr>
<td>Ventilation inhomogeneity</td>
<td>LCI&lt;sup&gt;f&lt;/sup&gt;</td>
<td>All ages</td>
<td></td>
</tr>
<tr>
<td>Pulmonary exacerbations*</td>
<td>Variable definitions</td>
<td>All ages</td>
<td></td>
</tr>
<tr>
<td>Nutritional status</td>
<td>Not specified</td>
<td>All ages</td>
<td></td>
</tr>
<tr>
<td>Clinical scoring tool: patient*</td>
<td>CFRSD-CRSS: CF-specific symptoms &amp; emotional and activity impact score</td>
<td>&gt;12 years</td>
<td></td>
</tr>
<tr>
<td>Combined clinical scoring tool</td>
<td>SK-m</td>
<td>All ages</td>
<td></td>
</tr>
<tr>
<td>Systemic inflammation</td>
<td>Change in C-reactive protein during exacerbation</td>
<td>All ages</td>
<td></td>
</tr>
</tbody>
</table>

*Outcomes that correlate with or capture how people feel, function or survive.

aFEV₁: Forced expiratory volume in 1 second.
bQoL: Quality of life.
cSTOP: Standardised Treatment of Pulmonary Exacerbations (STOP) study.
dCT: computerised tomography.
eCFQ: Cystic Fibrosis Questionnaire.
fMRI: magnetic resonance imaging.
gFEV<sub>0.5</sub>: Forced expiratory volume in 0.5 s; .hFEV<sub>25–75</sub> Forced expiratory flow between 25 and 50% of forced vital capacity (mid-expiratory flow); iSRAW: specific airway resistance; jLCI: Lung clearance index; kSK-m: Modified Schwachman-Kulczycki score; lFVC: forced vital capacity.

treatment burden and impact on school/work and relationships. There are four versions: two for children (interviewer format for 6–11 years and self-report format for 12–13 years), one for carers/parents (proxy report for children 6–13 years), and a teen/adult version (>14 years) [96–98].

Numerous radiological scores have been used to try to quantify the extent of structural lung damage using different imaging modalities (chest x-ray [CXR], computerised tomography [CT] and magnetic resonance imaging [MRI]) [see Table 2]. C-reactive protein (CRP) and white cell count were the most commonly reported biomarkers of inflammation [14,32,38,45,48,63,66,83,99–102].

Five studies reported costs associated with treatment for pulmonary exacerbations; these included direct costs (e.g. cost per pulmonary exacerbation within the hospital or via hospital-in-the-home therapy or annual cost associated with treatment of pulmonary exacerbations) and indirect costs (e.g. due to absence from work/school and loss of productivity or travelling expenses) [5,72,91,103,104].

3.2. Outcomes and their correlation with disease-related morbidity/mortality

Please see Table 3 for a summary of the reported correlation between reported outcomes with death and other measures [21,23,31,35,37,95,105–113].

While FEV₁ does not directly capture how people feel, function or survive, it has been validated as a surrogate measure of risk of
Table 4
Novel outcomes and endpoints proposed for evaluation in registered (unpublished) trials.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Endpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical Trials Register</td>
<td></td>
</tr>
<tr>
<td>NCT04354038 Gene expression and proteins and lung function</td>
<td>Changes in expression of genes and proteins over the course of treatment for pulmonary exacerbation</td>
</tr>
<tr>
<td>NCT04174664 Functional capacity Airway inflammation</td>
<td>Change in quadriceps fatigue and fatigue using visual analogue scale during exacerbation</td>
</tr>
<tr>
<td>NCT00684346 Airway obstruction</td>
<td>Proportion who achieve &gt; 90% of the baseline FEV₁ predicted value or the change in FEV₁ at 52 weeks</td>
</tr>
<tr>
<td>NCT04058548 Exercise capacity</td>
<td>Change in number of 1-minute STS® repetitions or 1 min STS® power; timed STS® repetitions x10 repetition and 6MWT® power; change in step count measured by pedometer</td>
</tr>
<tr>
<td>NCT01759342 Flexibility</td>
<td>Change in humeral distance, shoulder range of motion and hamstring length</td>
</tr>
<tr>
<td>NCT03497117 Lung ventilation</td>
<td>Change in percentage defect volumes</td>
</tr>
<tr>
<td>NCT02606487 Overall success score</td>
<td>Global outcome; unspecified</td>
</tr>
<tr>
<td>NCT04016571 Sleep quality</td>
<td>Change in time in bed, time asleep and time awake/restless measured by consumer wearable device</td>
</tr>
<tr>
<td>NCT01306279 Sputum microbiota</td>
<td>Change in relative abundance, dominance, evenness, diversity and richness between day 0, 5 and 14, altered constitution based on 16S® result</td>
</tr>
<tr>
<td>NCT02188758 Sputum P. aeruginosa</td>
<td>Change in P. aeruginosa gene expression post treatment and time to eradication over 108 weeks post treatment. Virulence gene determinants.</td>
</tr>
<tr>
<td>European Clinical Trials</td>
<td></td>
</tr>
<tr>
<td>2016–002.832–34 Hospitalisation</td>
<td>Rate of hospitalisation for respiratory event</td>
</tr>
<tr>
<td>2016–002.832–34 Sputum</td>
<td>Change in 16S sputum microbiome from day 0 to 14, culture conversion within 6 months from baseline</td>
</tr>
<tr>
<td>NCT01641822 Pulmonary exacerbations</td>
<td>Rate of pulmonary exacerbations from day 1 to week 24</td>
</tr>
<tr>
<td>2011–001.255–36 P. aeruginosa serology</td>
<td>Change in P. aeruginosa antibody titres</td>
</tr>
<tr>
<td>2011–001.255–36 Airway reactivity</td>
<td>Study-drug induced bronchospasm at day 1 and 28</td>
</tr>
</tbody>
</table>

n = 53.

a Directly reflect how patients feel, function or survive.
b 18FDG-PET: 18Fluoro-2-deoxy-d-glucose positron emission tomography-computed tomography; STS: Sit-to-Stand test; c 6MWT: 6-minute-walk test; d 16S: fungal PCR testing.

decision and reduced QoL, since low FEV₁ values (reflecting poor lung function) are strongly associated with increased risk of death and reduced quality of life [13,5–711–18,20–27,39,32,33,35–54,56–75,77–91,93–101,103–103,105,107,109,110,112–136]. The extent of structural lung damage quantified using CT scores was found to correlate with lung function and disease progression and was predictive of pulmonary exacerbations and death [106]; MRI scores were also found to correlate with the frequency of pulmonary exacerbations, with improvement in scores seen following receipt of therapy for pulmonary exacerbations [109,113].

The correlation between lung clearance index (LCI) and FEV₁ was variable, and while a moderate correlation with structural lung disease was shown, a direct correlation of LCI with risk of death has not been established. The CFRS-D-CRISS score was found to correlate only weakly with ppFEV₁ [31,37]. While patient-reported clinical scores and measures of QoL directly capture how people with CF feel and function, a correlation with risk of death has not been proven.

Pulmonary exacerbations have been associated with deteriorating lung function, with baseline lung function not recovered in one-third of episodes [35]; exacerbations have also been found to correlate with reduced QoL and decreased survival [31].

In cohort studies of people with CF, weight and lung function have been found to be significantly reduced in people with CF who die compared to those who don’t [21]. It is unclear however whether poor growth contributes to the risk of death independently of worsening lung function [137].

3.3. Evolution in reporting of trial outcomes and endpoints

The first CF clinical scoring system was the SK-score developed in 1958 [137,138]. The original and subsequently modified SK-scoring tool evaluated respiratory outcomes and radiological findings, however pulmonary function measures were not included [95]. The CFCS developed by [95] in 1999 included some pulmonary signs such as respiratory rate, decreased breath sounds or presence or absence of wheezeing or crackles, however it has not been used widely. We could find no evidence that the available scoring systems have been developed in collaboration with people affected by CF.

Work has increasingly been invested in identifying useful outcomes and endpoints for children with CF less than 6 years old. Specifically, there has been a focus on evaluating lung structure, pulmonary ventilation and perfusion using standardised scoring systems [please see Table S4].

Attention continues to be paid to investing in research designed to identify candidate biomarkers of inflammation in sputum and blood which could be used as substitutes for clinical endpoints [47].

A review of the Clinical Trials and European Clinical Trials registries [107] identified various novel outcomes and endpoints proposed for evaluation in registered but unpublished trials; these are listed in Table 4. Overall, the chosen outcomes and endpoints are increasingly diverse, with little sign of movement towards those most likely to be meaningful. Two outcomes not previously stud-
ied include an evaluation of the sputum microbiome and the genetic constitution of microorganisms present in the sputum.

4. Discussion

This study, the first systematic review of the range of outcomes and endpoints reported in studies of CF pulmonary exacerbations, found a wide range of outcomes and endpoints reported. Arguably, death, QoL and many patient-reported outcomes that capture how people with CF feel or function are likely to be intrinsically meaningful. Some additional measures, such as FEV₁ and the extent of structural lung damage, have been shown to correlate with death or other clinically meaningful outcomes. Outcomes that don’t capture how people feel, function and/or survive [139,140], and for which a correlation with risk of death or other clinically meaningful outcomes has not been established, are less likely to be meaningful; these include airway microbiology, markers of systemic inflammation, and ventilation inhomogeneity.

While there is increasing recognition of the value of involving people with CF when choosing trial outcomes and endpoints, this is not currently mandated by regulators, nor is there consensus in the best approach for doing this [4]. Regarding management of pulmonary exacerbations, the Standardised Treatment of Pulmonary Exacerbations (STOP) trial group have identified several outcomes of interest to CF clinicians [4]. While outcomes of interest to people with CF have also been identified [37,141], we found no evidence that people with CF had influenced the choice of outcomes used in trials of CF pulmonary exacerbations published to date [142].

A change in FEV₁ (absolute value or ppFEV₁) between two time points has been the most commonly used primary endpoint, and is still the only endpoint accepted by the European Medicines Agency and the US Food and Drug Administration [143] since it has been shown to correlate strongly with increased risk of death and reduced quality of life [4]. We found some limitations to its use, including that it cannot be reliably performed in young children, precluding its use in trials of preschool children and infants. When FEV₁ is chosen as an endpoint, it is unclear what is the most appropriate comparison to make in order to capture the effect of the treatment. The STOP [31] trial group have warned that comparing a post-treatment FEV₁ to a ‘previous best’ FEV₁ result may be problematic, as this data may be difficult to locate, resulting in missing data points [35]. Further, FEV₁ is well preserved in early disease and exhibits reduced variability in end-stage disease, reducing its ability to capture changes in lung function in these groups [35].

QoL has been widely used as an outcome in CF pulmonary exacerbation trials. QoL is a multifaceted construct; while some QoL tools have been developed in conjunction with consumers, some QoL scores may not reliably or accurately capture QoL [40]. Disease-specific QoL measures may be more sensitive to changes in health state and provide additional clinically meaningful information than generic QoL measures [121]. Differences may occur in how younger and older people with CF respond to QoL questionnaires because symptoms differ by disease stage, and the importance placed on specific health outcomes may also vary by age and disease stage. Differences in how children and their carers perceive their QoL have also been described; this should be considered when deciding whether outcomes should be reported from the perspective of the child or their proxy [144]. The validity, responsiveness and reliability of different QoL instruments is variable; this subject is beyond the scope of this review and will be reported separately.

While candidate sputum and blood biomarkers of inflammation have been reported in trials of CF pulmonary exacerbations, none have been shown to be desirable endpoints; desirable criteria include being reproducible, feasible, and sensitive and specific to treatment effects (ideally closely related to the causal mechanisms of the disease outcomes). Nor have they been shown to correlate with risk of death or other meaningful clinical outcomes [96].

Clinical trials should choose outcomes that help address the trial objective(s) and which have been identified by people with CF as meaningful. The choice of outcomes may also be influenced by the population under study, the trial setting, available expertise and equipment, and cost. Selection may also be limited by the availability of tools to reliably measure the outcome(s) of interest [101].

A limitation of this review is that the relevant data on outcomes and endpoints from individual trials was often poorly reported in reviews and systematic reviews. Because we aimed to describe the range of outcomes and endpoints which have been reported in CF studies regardless of their quality, we did not undertake an assessment of the quality of those studies.

This review is a first step towards an agreed COS for trials of pulmonary exacerbations in people with CF. Outcomes that capture (directly or indirectly) how people feel, function or survive are more likely to be meaningful to people with CF than those which do not, and are highlighted here. It will be necessary to involve people with CF and their families to confirm the set of outcomes and endpoints of most importance to them; it is not clear that they have been widely consulted by investigators before now. Specific attention is required to identify suitable outcomes and endpoints for children less than six years old. It is hoped that a COS will guide end-users involved in the design, conduct, reporting and translation of research findings to inform best practice and ultimately improve outcomes for people living with CF.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors. CM is supported by an NHMRC post-graduate scholarship (GNT1150996), and top-up grants from the Wesfarmers Centre (top-up grant) and the Perth Children’s Hospital Foundation (9772 & 9757). TS is supported by a Career Development Fellowship from the National Health and Medical Research Council (GNT1111657). CB is also supported by the National Health and Medical Research Council (1111596/1173163).

Authors’ contributions

CM was responsible for the study conceptualisation, data curation and overall methodology. CM and JW were responsible for article selection. TS, CM, SS, SW and Andre Schultz and Alan Smyth elaborated the study protocol. CM drafted the manuscript. All authors were involved in the interpretation of data and revision of the manuscript. All authors approved the final manuscript.

Declaration of Competing Interest

Nil conflicts to declare.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jcf.2020.08.015.

References


et Al-Aloul
Curtis
Dentice.

Water's and Renners


Waters and Renners


Waters and Renners


Waters and Renners


Waters and Renners


Waters and Renners


Waters and Renners


Waters and Renners


Chapter 5: Measurement properties of tests and tools used to capture outcomes in CF studies

5.1 Chapter summary

In this chapter, I systematically review and summarise the characteristics and measurement properties of tests and tools which have been used to capture outcomes in studies among people with CF, including their reliability, validity and responsiveness. This review is intended as a guide for researchers when selecting tests or tools used to capture outcomes to generate evidence for measuring treatment effects in CF trials.

There is no existing consensus for the selection of tests and tools for measuring outcomes in trials in people with CF. Results of this chapter represent a first step towards development of consensus methods for measurement of core outcomes selected for inclusion in a core outcome set (COS). This will be a focus for post-doctoral studies (see Chapter 9).

5.2 Journal article


Appendix 1: COSMIN Definitions for measurement properties
Appendix 2: Abbreviations and their meanings
S1: Search strategy (Medline)
S2: Search strategy (Embase)
S3: Studies that met inclusion criteria
S5: Characteristics of tests and tools
S6: Measurement properties of tests and tools

Appendices and supplementary materials are available here.
The measurement properties of tests and tools used in cystic fibrosis studies: a systematic review


1Wesfarmers Centre for Vaccines and Infectious Diseases, Telethon Kids Institute, Nedlands, Australia. 2Division of Paediatrics, Faculty of Medicine, University of Western Australia, Nedlands, Australia. 3Infectious Diseases Dept, Perth Children's Hospital, Nedlands, Australia. 4Physiotherapy Dept, Sir Charles Gairdner Hospital, Nedlands, Australia. 5Sydney School of Public Health, The University of Sydney, Sydney, Australia. 6Centre for Kidney Research, The Children's Hospital at Westmead, Westmead, Australia. 7Wal-yan Respiratory Research Centre, Telethon Kids Institute, University of Western Australia, Nedlands, Australia. 8Dept of Respiratory and Sleep Medicine, Perth Children's Hospital, Nedlands, Australia. 9School of Population Health, Curtin University, Bentley, Australia. 10Evidence Based Child Health Group, University of Nottingham, Queens Medical Centre, Nottingham, UK. 11Pathwest Laboratory Medicine WA, QEII Medical Centre, Nedlands, Australia. 12St John of God Hospital, Subiaco, Australia. 13School of Population Health and Preventive Medicine, Monash University, Melbourne, Australia. 14Menzies School of Health Research, Royal Darwin Hospital Campus, Tiwi, Australia.

Corresponding author: Charlie McLeod (charlie.mcleod@health.wa.gov.au)

Abstract
There is no consensus on how best to measure responses to interventions among children and adults with cystic fibrosis (CF). We have systematically reviewed and summarised the characteristics and measurement properties of tests and tools that have been used to capture outcomes in studies among people with CF, including their reliability, validity and responsiveness. This review is intended to guide researchers when selecting tests or tools for measuring treatment effects in CF trials. A consensus set of these tests and tools could improve consistency in how outcomes are captured and thereby facilitate comparisons and synthesis of evidence across studies.

Introduction
Research is conducted to generate evidence of the efficacy and safety of interventions to inform best clinical practice and thereby improve outcomes for patients. When designing studies, it is necessary to establish which outcomes must be evaluated to meet the study objectives and how these outcomes will be measured and analysed as end-points [1]. Tests or tools may be required for outcome measurement. To improve consistency and facilitate synthesis of evidence across studies, there is a need to establish a consensus set of these tests and tools for measuring outcomes in studies in people with cystic fibrosis (CF). These must be responsive to changes in the outcome of interest and capture outcomes with sufficient validity, reliability and precision [2, 3]. This is necessary so that results can be interpreted with confidence and be used to support the translation of evidence into practice.

There is no existing consensus on the selection of tests or tools for measuring outcomes. Selection is challenging for a number of reasons. First, tests and tools may lack appropriate validation, and hence their quality might be uncertain. Secondly, criteria to facilitate interpretation of the results of the test or tool may not exist. Thirdly, logistic or economic constraints may restrict the use of some. Finally, although initiatives to improve and standardise the use of patient-reported outcome measures (PROMs) have been
established [4, 5], there is no standardised approach for evaluating and selecting optimal tests and tools more generally in clinical research.

As a first step towards developing a consensus set of tests and tools for measuring outcomes in CF studies, we aimed to evaluate and summarise the characteristics and properties of tests and tools that have been used in previous CF studies. In the interim, we hope this review will be used by clinicians, people with CF, researchers and policy makers to guide optimal selection of these tests and tools, and to encourage validation or development of new tests or tools for measurement where required.

Methods

Search strategy and selection criteria

This was a PROSPERO registered systematic review (CRD42020151785). The search strategy is provided in the supplementary material. Medline, Embase and the Cochrane database were searched from inception until July 2020. Outcome measures proposed for evaluation in the Clinicaltrials.gov registry were also evaluated.

Inclusion criteria were reports of randomised controlled trials, observational studies, conference abstracts and reviews written in English, evaluating one or more measurement properties of a test or tool used to measure health outcomes in studies among people with CF. Original studies were sought to provide additional information about the characteristics or measurement properties of the tests and tools where necessary. Registered trials without published results proposing evaluation of one or more measurement properties of novel tests or tools were also included. Exclusion criteria were tests or tools developed for diagnostic purposes or used for evaluation of microbiological outcomes, or validation studies written in languages other than English. Tests and tools used in people with CF but validated in non-CF populations were beyond the scope of this review.

Titles and abstracts were screened independently by two reviewers (C. McLeod and J. Wood). Potentially eligible studies were downloaded to Endnote and duplicates removed. Full text articles were retrieved and eligibility confirmed by consensus of the reviewers. A third reviewer (T.L. Snelling) was used to confirm eligibility where consensus could not be achieved. Relevant data were extracted by C. McLeod and recorded in an Excel database and cross-checked by J. Wood.

The following characteristics of selected tests or tools were recorded: the outcome construct measured; the target population; mode of administration of the test or tool; recall period (if relevant); time taken to perform the test; the range of possible scores; and the ease of use (feasibility). The following properties of measurement were critically appraised: 1) validity, including content validity, construct validity (including convergent and discriminant performance of the test, the structural validity and cross-cultural validity) and criterion validity (including concurrent and predictive validity); 2) reliability, including the test–retest and inter-/intra-rate reliability, internal consistency and measurement error; 3) responsiveness; and 4) the minimal clinically important difference (MCID). Definitions for these measurement properties were based on those provided by the Consensus-based Standards for the selection of health Measurement INstruments initiative (COSMIN); these are presented in supplementary appendix 1.

Definitions, abbreviations and citations

Quality of life (QoL) tools were broadly defined as those which capture an individuals’ perception of their life satisfaction relative to their goals in the context of their culture and value systems, and not those that capture QoL based solely on the health status of the individual per se (health-related quality of life; HRQoL) [6]. Disease-specific QoL tools were defined as those developed for measuring QoL in people with CF, whereas generic QoL tools were defined as those originally developed for use in other populations that have also been applied in studies involving people with CF.

A full list of abbreviations and their meanings used throughout this manuscript and supplementary materials are alphabetically listed in supplementary appendix 2. References for information presented in the tables throughout this manuscript are provided in the supplementary materials.

Results

The review process is depicted in figure 1. 118 studies evaluating the measurement properties of 74 tests and tools used in studies among people with CF identified from Medline, Embase or Clinical Trials met the inclusion criteria [7–119]; a summary of these studies is provided in table S3. This review included three registered studies proposing validation of tests or tools used in people with CF with unpublished
results [120]. Nine source articles describing the characteristics or measurement properties of tests or tools were also included [121–129].

Characteristics of tests and tools
Tests or tools were categorised as PROMs capturing QoL or other patient-reported outcomes, clinical scores, radiological scores or tests capturing functional exercise performance, CF transmembrane conductance regulator (CFTR) function or sputum characteristics.

QoL tools
17 generic QoL tools evaluated in CF populations and seven CF-specific QoL were identified. Characteristics of these generic and CF-specific QoL tools are detailed in table S5 and table 1, respectively.

The Cystic Fibrosis Questionnaire-Revised (CFQ-R; original version 2003) has been the most widely used QoL tool reported in CF studies and is available in 34 languages [130]. It is endorsed for use in clinical trials by the US Food and Drug Administration (FDA) and European Medicines Agency [131, 132]. There are five versions available; these are described in table 1.

One QoL questionnaire for use by carers of people with CF was identified, the Carer QoL in CF questionnaire (CQOLCF); this is a 35-item questionnaire designed to evaluate how providing care for someone with CF impacts on a carers’ physical, emotional and social functioning and family [69].

Tools capturing patient-reported outcomes (excluding QoL)
Six questionnaires designed to evaluate self-reported levels of physical activity were identified [86, 101]. Two questionnaires capturing body image for use in people aged ≥14 years [116] and one tool measuring dietary intake in children aged between 7 months and 12 years (table S5) were also found [40]. One PROM has been used to evaluate the impact of CF on stigma, disclosure, public attitudes and negative self-image among adults with CF and their carers [96]. A separate PROM originally developed for use in people with asthma has been used to capture work productivity and activity in people with CF [92].

12 clinical scores calculated from outcome data reported by people with CF were identified; three were developed for use in CF pulmonary exacerbations [12, 106], three captured respiratory symptoms [22, 74,
<table>
<thead>
<tr>
<th>Test or tool</th>
<th>Description</th>
<th>Constructs(s)</th>
<th>Target population</th>
<th>Administration</th>
<th>Recall period</th>
<th>Range of scores</th>
<th>Feasibility/ cultural validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>CFIQ</td>
<td>40 items; 5–6 min to complete</td>
<td>Activity limitation (physical, social, leisure), school/ work limitations, vulnerability/lack of control, emotional impact, treatment burden and future outlook</td>
<td>Children and adults with CF and their carers; interview templates for children aged 6–11 years, adolescents, adults &gt;12 years and carers of children aged 0–18 years</td>
<td>Paper</td>
<td>36 items 7 day recall, the remainder “current status”</td>
<td>5- or 7-point verbal rating scale</td>
<td>Largely developed in Caucasian population; further validation required</td>
</tr>
<tr>
<td>CFQoL</td>
<td>QoL and symptom scoring tool; 52 items over nine domains; 15–20 min to complete</td>
<td>Two symptom scales (chest and emotional) and seven QoL domains: physical functioning (10), social functioning (4), treatment issues (3), future concerns (6), interpersonal relationships (10), body image (3), career issues (4)</td>
<td>Adults and adolescents</td>
<td>Paper</td>
<td>14 days</td>
<td>Each response 6-point Likert scale; total score 0–100</td>
<td>Time consuming</td>
</tr>
<tr>
<td>CFQoL scale (single item)</td>
<td>VAS: how has CF affected your QoL in the last 2 weeks? Couple of mins to complete</td>
<td>Single QoL question</td>
<td>Adults</td>
<td>Paper</td>
<td>14 days</td>
<td>0–10</td>
<td>Quick validated in population who had not had pulmonary exacerbation for past 6 weeks</td>
</tr>
<tr>
<td>CFQ-R</td>
<td>QoL and symptom scoring tool; four scales: 1) CFQ-R &gt;14 years (44 items over 9 domains); 2) CFQ-R child (8 domains, 35 items), interviewer-administered 6–13 years and self-report 12–13 years; 3) CFQ-R parent: 44 items, 10 min; 4) preschool 3–6 years, 28 items, 15 min to complete</td>
<td>Activity limitation (physical, social, leisure), school/ work limitations, vulnerability, lack of control, emotional impact, treatment burden and future outlook</td>
<td>CFQ-R &gt;14 years; CFQ-R child: interviewer-administered 6–13 years and self-report 12–13 years CFQ-R parent: proxy report for children 6–13 years Separate are not proxy report for children 4–60 months</td>
<td>Paper/electronic</td>
<td>14 days</td>
<td>4-point Likert scale; total score 0–100</td>
<td>Most widely used HRQoL questionnaire in CF; translated into 34 languages; EMA/FDA supports use in clinical trials</td>
</tr>
</tbody>
</table>
three characterised pain [90, 133], two quantified abdominal symptoms [28, 31, 41] and one has been used to evaluate physical and psychological symptom burden [105, 106] (table 2). Of these, the CFRSD-CRISS (chronic respiratory infection symptom score) has been the most widely used in CF studies and is available in 38 languages; it evaluates eight respiratory symptoms and is validated for use in people with CF aged >12 year [44].

### Clinical scores

The modified Schwachman scale (first described in 1964) was developed as a longitudinal clinical assessment tool. It includes activity levels, chest findings, cough, growth, nutrition, the character of stool and radiological changes; lung function is not included in this measure [26]. A test developed by RADINE et al. [45] measuring nocturnal cough over two consecutive nights was found to be safe and feasible.

Three prognostic scoring tools were found. The most recent, in 2004, was a 5-year survival prediction tool [63], which was designed to guide eligibility for lung transplant; survival is predicted based on age, sex, forced expiratory volume in 1 s (FEV1 % pred), weight for age z-score, pancreatic function, presence of diabetes, infection with *Staphylococcus aureus* or *Burkholderia cepacia* and the annual number of pulmonary exacerbations. The second, the modified Huang score (first described in 1997) [26], was developed for use as a prognostic and longitudinal assessment tool for those with terminal disease and captures clinical features (including lung function), radiological features and complications of disease. The oldest tool, first described in 1973, is the National Institutes of Health score (NIH) [26, 108], which was developed for people aged 5–30 years. It predicts the probability of death within 3 years based on lung function, chest radiograph (CXR) changes, and physiological and psychological features.

---

**TABLE 1 Continued**

<table>
<thead>
<tr>
<th>Test or tool</th>
<th>Description</th>
<th>Constructs(s)</th>
<th>Target population</th>
<th>Administration</th>
<th>Recall period</th>
<th>Range of scores</th>
<th>Feasibility/ cultural validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>eCF-QUEST</td>
<td>Electronic, 3 domains, 4 items</td>
<td>Global measure (40 items), gastrointestinal (5 items) and general health (2 items)</td>
<td>Adults</td>
<td>Paper/electronic</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>DISABKIDS-CF</td>
<td>6 items, 2 min to complete</td>
<td>Impact and treatment dimensions</td>
<td>Exclusively for use in children and adolescents 8–17 years; self-report and proxy version (carer)</td>
<td>Paper</td>
<td>Each dimension 0–100</td>
<td>5-point Likert scale; scores 0–100% (higher score=higher QoL)</td>
<td>English and Spanish versions</td>
</tr>
<tr>
<td>FLZ-CF</td>
<td>Healthy and general patient population; 18 items over 8 domains, 9-item weighted scale, 5 min to complete</td>
<td>9 items: cough/dyspnoea, abdominal, sleep, eating, therapy routine, adherence, understanding by others, being needed by others, disadvantage</td>
<td>Adolescents &gt;15 years and adults</td>
<td>Paper</td>
<td>28 days</td>
<td>0–100 (high score=high satisfaction)</td>
<td>Screening test</td>
</tr>
<tr>
<td>CQOLCF</td>
<td>35 item carer QoL instrument, &lt;10 minutes</td>
<td>Physical, emotional, family and social functioning</td>
<td>Carers of people with CF</td>
<td>Paper</td>
<td>NR</td>
<td>Each response 5-point Likert scale</td>
<td>NR</td>
</tr>
</tbody>
</table>

CFIQ: CF impact questionnaire; CFQoL: CF QoL; CFQ-R: revised CF questionnaire; eCF-QUEST: electronic version of the CFQoL evaluative self-administered test; FLZ-CF: Questions of Life Satisfaction; CQOLCF: caregiver QoL CF scale; VAS: visual analogue scale; HRQoL: health-related QoL; EMA: European Medicines Association; FDA: US Food and Drug Administration; NR: not reported.
<table>
<thead>
<tr>
<th>Test or tool</th>
<th>Description</th>
<th>Constructs(s)</th>
<th>Target population</th>
<th>Administration</th>
<th>Recall period</th>
<th>Range of scores</th>
<th>Feasibility/cultural validity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pulmonary exacerbations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AWESCORE</td>
<td>5 domains, each with 2 items</td>
<td>Respiratory (cough and sputum), physical (energy and exercise), nutritional (appetite and weight), psychological (mood and anxiety) and general health “wellness” and sleep</td>
<td>Adults</td>
<td>Paper</td>
<td>NR</td>
<td>0–100</td>
<td>NR</td>
</tr>
<tr>
<td>CFRSD/CFRSD-CRISS</td>
<td>Symptom score: respiratory and emotional items (respiratory only in short version); developed for pulmonary exacerbation</td>
<td>8 respiratory symptoms, 4 emotional items and 4 other items (or short-version CFRSD-CRISS; 8 respiratory items: difficulty breathing, fever, tired, mucus, chills/sweats, cough, mucus, chest tightness, wheezing)</td>
<td>&gt;12 years</td>
<td>Paper/electronic</td>
<td>Daily</td>
<td>3–4 point Likert scale, total score 0–100</td>
<td>Available in 38 languages</td>
</tr>
<tr>
<td><strong>Symptom score system</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Respiratory</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ReS-CF</td>
<td>4-item questionnaire; &lt;1 min to complete</td>
<td>Self-reported VAS for respiratory symptoms, cough, chest congestion and sputum</td>
<td>Adults</td>
<td>Paper</td>
<td>NR</td>
<td>Each VAS scored separately 0–10 (worst)</td>
<td>Screening tool: respiratory symptoms</td>
</tr>
<tr>
<td>SOBQ</td>
<td>Developed for PEX assessment; 17 items (13 respiratory and 4 CF-related impacts)</td>
<td>0–6 years and 7–11 years</td>
<td>NR</td>
<td>NR</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOBQ</td>
<td>24 items; patients with COPD, CF and lung transplant recipients</td>
<td>Measures SOB while performing activities of daily living</td>
<td>Adults</td>
<td>Paper</td>
<td>NR</td>
<td>5-point Likert scale for each response; scores 0–120 (worst)</td>
<td></td>
</tr>
<tr>
<td><strong>Pain</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BPI</td>
<td>Severity and impact of pain on daily functions in people with chronic diseases; short: 5 min, long: 10 min to complete</td>
<td>7 domains: general activity, mood, walking ability, normal work (including housework), relationships with others, sleep, enjoyment in life</td>
<td>Adults</td>
<td>Paper</td>
<td>Daily</td>
<td>NR</td>
<td>Psychometrically and linguistically validated in 24 languages</td>
</tr>
<tr>
<td>DPAQ-CF</td>
<td>7 items</td>
<td>Frequency, duration, intensity, location and coping response to pain</td>
<td>Adolescents and adults</td>
<td>Paper/electronic</td>
<td>Daily</td>
<td>5-point Likert scale for each response; total score 0–10</td>
<td></td>
</tr>
</tbody>
</table>

Continued
Lung function tests

Tests used to measure lung function for which measurement properties were described include spirometry, raised volume rapid thoracic compression (RVRTC), impulse oscillometry and lung clearance index (LCI). Characteristics of these tests are reported in table S5.

Spirometry has been the most frequently used lung function test in CF studies, and the measure of lung function most commonly reported has been FEV₁ [30]. This has been variously measured as the crude volume (in litres) or as the percentage predicted volume for age and height, or z-score for age, sex, height and ethnicity; within-individual changes in the FEV₁ have been reported as either the absolute change, or change relative to the baseline measure [54, 72].

Imaging scoring tools

Four CXR scoring systems used to quantify the degree of structural lung disease were identified. The oldest, the Chrispin–Norman score in 1974, is based on chest configuration and the presence or absence of different types of “shadows” [123]. First described in 1979, the Brasfield or Birmingham score (scored between 3 and 25) aims to capture radiographic evidence of air trapping, bronchial wall thickening, bronchiectasis, atelectasis and general severity [73, 122]. From 1993, the Wisconsin score (0–100) has been used to evaluate six attributes including hyperinflation, peribronchial thickening, bronchiectasis, opacities and atelectasis [54, 129]. The Brasfield scoring system has been reported to be easier to use and quicker to perform than the Wisconsin score [73]. The Northern score (introduced in 1994) is calculated based on the presence of linear, cystic or confluent opacities in each lung quadrant rated on a four-point Likert scale (normal to very severe) on a single film [124].

Three computed tomography (CT) scoring tools were identified: the Brody score (I and II), originally developed in 1999, and the CF-CT score (introduced in 2011), which is based on the Brody II score and...
aims to improve standardisation of the latter [111]. These tools have been used in people aged >5 years. In 2014, the Perth Rotterdam Annotated Grid Morphometric Analysis method (PRAGMA-CT) score was developed for application in children and infants [102] and takes an experienced person ∼30 min to calculate per CT scan [111].

Scoring tools based on quantitative magnetic resonance imaging (MRI) are still in development [54].

**Functional exercise performance**
The most frequently studied field exercise test performed in people with CF is the 6-min walk test (6MWT) [36]. Characteristics of tests used to measure functional exercise performance are summarised in table 3.

**CFTR function**
Characteristics of tests used to directly (e.g. sweat chloride tests) or indirectly (e.g. nasal potential difference tests) measure CFTR function are summarised in supplementary table S5.

**Sputum tests**
Rheometry tests which capture the characteristics of sputum, and tests used to capture markers of inflammation in sputum, are summarised in table S5.

**Measurement properties of tests, tools or instruments**

**QoL tools**
The measurement properties of generic QoL tools based on their evaluation in CF populations are detailed in table S6, and the properties of CF-specific QoL tools are summarised in table 4.

The development of the CFQ-R involved people with CF, and it has been shown to be reliable, with sound content (including face) validity. The tool has been shown to have good internal consistency for all constructs examined, including parent proxy report of physical, eating and respiratory subscales (α=0.73–0.86), but not for treatment burden (r=0.44). The CFQ-R score correlates with FEV1 and body mass index, and discriminates different degrees of disease severity [130], but a correlation with mortality has not been reported. Based on clinician judgement, a change of 0.8 units in the CFQ-R score was considered the MCID in the context of treatment for pulmonary exacerbations [134].

**Patient-reported symptoms and function**
The measurement properties of patient-reported symptoms and function are summarised in table 5.

**Clinical scores**
A validation study that evaluated nocturnal cough as an outcome found people with CF coughed more than healthy subjects (p<0.001); the reliability for repeated measurements was higher when cough epochs were scored (multiple coughs with <2 s between individual coughs) compared to discrete coughs (internal consistency coefficient (ICC) 0.75 versus 0.49, respectively) [45].

The interobserver reliability of the modified Schwachman score captured as Pearson’s r coefficient was 0.71, 0.64 and 0.85 for the history, examination and growth domains, respectively [26]; the correlation was 0.92 with the NIH score and 0.67 with FEV1.

The internal consistency of the modified Huang score was reported to be α=0.6 (except the domain relating to complications). The correlation of this score with FEV1 % pred in moderate (score 35–60, r2=0.3) and severe disease (<35 points, r2=0.43) was greater than in asymptomatic or mild disease [26]. The NIH score was found to be significantly lower in the 5 years before death compared to CF controls who did not die (p=0.001); those with a score between 61 and 70 had a 25% chance of dying within 3 years. The internal consistency of this score was reported to be α=0.81 and the inter-rater (Pearson’s r) score was 0.90; this was predominantly attributed to the robustness of the pulmonary domain on subscale analysis [26].

**Lung function tests**
Low FEV1 was shown to correlate with death, with a relative risk of death within 2 years of 2.0 (95% CI 1.9–2.2) for each 10% reduction in FEV1 below the predicted value after adjustment for age and sex [126]. Among people with the same FEV1, the risk of death was more than double for females compared to males (RR 2.2 (95% CI 1.6–3.1)). FEV1 was also shown to correlate with QoL; a 5% change in FEV1 was associated with a change in CFQ-R score from 0.5 to 2.3 points [125].
The RVRTC test demonstrated good test–retest reliability with a coefficient of variation reported to be 2–6%; it differentiated people with CF from healthy controls, including among children aged <6 years [91]. Parameters were shown to improve in children aged between 4 months and 1 year, raising the possibility that lung damage may be reversible during this time [91]. However, RVRTC testing has not been appropriately standardised and consequently has not yet been recommended by authoritative bodies such as the European CF Society Clinical Trial Network as a primary outcome measure for use in CF studies [32].

Measurement of LCI has been found to be reliable, valid and responsive during treatment of pulmonary exacerbations and for monitoring disease progression [23, 37, 89, 110]. Measurements obtained by N₂ washout and by SF₆ were comparable (limits of agreement −2.5 to 1.2) [23]. These tests were found to

<table>
<thead>
<tr>
<th>TABLE 3</th>
<th>Functional measures of exercise capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test or tool</td>
<td>Description</td>
</tr>
<tr>
<td>iSTEP</td>
<td>Externally paced test; speed increases every 2 min</td>
</tr>
<tr>
<td>MSWT</td>
<td>15-level modification of ISWT Office based walk/run test</td>
</tr>
<tr>
<td>PowerSTS</td>
<td>1-min sit-to-stand power index</td>
</tr>
<tr>
<td>Triple hop distance</td>
<td>Starting at one end of a tape, asked to hop three times consecutively on dominant leg, trying to cover as much distance as possible</td>
</tr>
<tr>
<td>Vertical jump test</td>
<td>90-cm² mat connected to a timer next to a wall; time off mat converted to a vertical jump (cm) using a controlled (90 degree) and uncontrolled knee angle</td>
</tr>
<tr>
<td>3MST</td>
<td>Submaximal stress test (distance covered in m)</td>
</tr>
<tr>
<td>6MWT</td>
<td>Submaximal stress test (distance covered in m)</td>
</tr>
<tr>
<td>30 s or 1-min-STS</td>
<td>Cardiorespiratory response during a 30-s or 1-min STS test (chair height 40 cm without armrest; full knee extension); as many repetitions as possible in 1 min</td>
</tr>
</tbody>
</table>

iSTEP: incremental field step test; MSWT: modified shuttle walk test; PowerSTS: 1-min sit-to-stand power index; 3MST: 3-min sit-to-stand test; 6MWT: 6-min walk test; STS: sit-to-stand test; ISWT: incremental shuttle walk test; N/A: not applicable; CF: cystic fibrosis; CLD: chronic lung disease.
<table>
<thead>
<tr>
<th>Test or tool</th>
<th>Content validity</th>
<th>Convergent validity</th>
<th>Discriminant validity</th>
<th>Concurrent</th>
<th>Predictive</th>
<th>Intra- or inter-rater and test-retest</th>
<th>Internal consistency</th>
<th>Measurement error</th>
<th>Responsiveness</th>
<th>Comments/ MCID</th>
</tr>
</thead>
<tbody>
<tr>
<td>CFIQ</td>
<td>Demonstrated; people &gt;6 years with CF and carers used in construction</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>Requires further validation; content validity established</td>
</tr>
<tr>
<td>CFQoL</td>
<td>Easy to understand/complete; people with CF involved in construction</td>
<td>Correlation of emotional scores with SF-36 r=0.64; p=0.001</td>
<td>Chest and emotional scores distinguished between severity of chronic lung disease (FEV1 % pred &gt;70, 40–70 or &lt;40)</td>
<td>Chest score correlation with FEV1 not tested</td>
<td>NR</td>
<td>Test–retest rs=0.74–0.94 (p=0.01) Robust after 7–10 days; 0.9 for emotional scores and 0.93 for respiratory</td>
<td>Cronbach’s α=0.3</td>
<td>NR</td>
<td>NR</td>
<td>Chest symptom scores increased during pulmonary exacerbation treatment Chest and emotional score responsive over a 2-week application period in hospital (47–70.3, p=0.006) versus home groups (49.7–68.8, p=0.03)</td>
</tr>
<tr>
<td>Test or tool</td>
<td>Content validity</td>
<td>Convergent validity</td>
<td>Discriminant validity</td>
<td>Concurrent validity</td>
<td>Predictive validity</td>
<td>Intra- or inter-rater reliability</td>
<td>Internal consistency</td>
<td>Measurement error</td>
<td>Responsiveness</td>
<td>Comments/MCID</td>
</tr>
<tr>
<td>-------------</td>
<td>------------------</td>
<td>---------------------</td>
<td>-----------------------</td>
<td>---------------------</td>
<td>-------------------</td>
<td>----------------------</td>
<td>---------------------</td>
<td>-----------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>CFQ-R</td>
<td>Patients involved in testing clarity of items Preschool version: children able to understand and answer questions</td>
<td>Correlation between CFQ-R and SF-36 on physical health perceptions/general health (r=0.81, p&lt;0.01), vitality (r=0.84, p&lt;0.01), role/role-physical (r=0.73, p&lt;0.01), emotional functioning/mental health (r=0.74, p&lt;0.01) and social domains Strong convergence between child and parent proxy reports, although children generally reported better HRQoL than parents</td>
<td>CFQ-R: no significant difference between age groups (6–11 years, 12–13 years versus ≥14 years) for all domains except treatment between 6– to 11-year-olds and ≥14-year-olds Significant association between CFTR genotype and CFQ-R scores (K=9.34, p&lt;0.01) Strong parent–child agreement found for scales measuring respiratory symptoms, but children reported more fatigue and difficulty running/walking Respiratory score established using FEV1; correlation with FEV1, %pred r=0.42, p-value NR; correlation with number of intravenous antibiotic courses r=−0.27, p-value NR</td>
<td>Respiratory score</td>
<td>NR</td>
<td>Acceptable</td>
<td>Cronbach’s α=0.6–0.76 with the exception of treatment burden (α=0.44) Parent proxy report for CFQ-R physical, eating and respiratory subscales α=0.73–0.86</td>
<td>NR</td>
<td>Based on clinician judgement, a moderate change=0.5 units and an important change=0.8 units pre- &amp; post-exacerbation</td>
<td>NR</td>
</tr>
<tr>
<td>Test or tool</td>
<td>Content validity</td>
<td>Convergent validity</td>
<td>Discriminant validity</td>
<td>Concurrent</td>
<td>Predictive</td>
<td>Intra- or inter-rater and test-retest</td>
<td>Internal consistency</td>
<td>Measurement error</td>
<td>Responsiveness</td>
<td>Comments/ MCID</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------------</td>
<td>---------------------</td>
<td>-----------------------</td>
<td>------------</td>
<td>-----------</td>
<td>-----------------------------</td>
<td>---------------------</td>
<td>-----------------</td>
<td>---------------</td>
<td>----------------</td>
</tr>
<tr>
<td>CF-QUEST</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>rₛ was 0.951 for the total CF-QUEST score, 0.929 for gastrointestinal module and 0.941 for GHQ module for paper/electronic versions</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>Excellent correlation and agreement of electronic version with its validated paper counterpart</td>
</tr>
<tr>
<td>DISABKIDS-CFM</td>
<td>NR</td>
<td>Convergent validity with KINDL-R established; r=0.6</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>Cronbach’s α=0.55 (p=0.011) for the impact dimension and 0.480 (p=0.02) for the treatment dimension</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Test or tool</td>
<td>Content validity</td>
<td>Convergent validity</td>
<td>Discriminant validity</td>
<td>Concurrent validity</td>
<td>Predictive validity</td>
<td>Intra- or inter-rater and test-retest</td>
<td>Internal consistency</td>
<td>Measurement error</td>
<td>Responsiveness</td>
<td>Comments/MCID</td>
</tr>
<tr>
<td>-------------</td>
<td>------------------</td>
<td>---------------------</td>
<td>-----------------------</td>
<td>---------------------</td>
<td>--------------------</td>
<td>---------------------</td>
<td>-------------------</td>
<td>------------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>FLZ-CF</td>
<td>NR</td>
<td>Pearson’s correlation r=0.75 with the generic satisfaction with health scale of the FLZ-CF, r=0.3 with FEV1 % pred and r=−0.26 with daily time for home therapy Leisure time/hobbies, physical condition, ability to relax, energy for life and satisfaction with health r_s &gt;0.5 with positive mood and ability to relax and SF-36 physical functioning, general health, vitality, social function and mental health</td>
<td>The scale discriminated significantly between patients with and without need for assistance with daily life and between patients with and without a partner</td>
<td>Physical condition/fitness and FEV1 % pred r_s=0.66</td>
<td>NR</td>
<td>NR</td>
<td>Cronbach’s α=0.82–0.89</td>
<td>NR</td>
<td>NR</td>
<td>Reliable and valid</td>
</tr>
<tr>
<td>Test or tool</td>
<td>Content validity</td>
<td>Convergent validity</td>
<td>Discriminant validity</td>
<td>Concurrent validity</td>
<td>Predictive validity</td>
<td>Intra- or inter-rater and test-retest</td>
<td>Internal consistency</td>
<td>Measurement error</td>
<td>Responsiveness</td>
<td>Comments/MCID</td>
</tr>
<tr>
<td>-------------</td>
<td>------------------</td>
<td>---------------------</td>
<td>-----------------------</td>
<td>---------------------</td>
<td>--------------------</td>
<td>---------------------</td>
<td>-------------------------------</td>
<td>------------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Single item CFQoL questionnaire</td>
<td>NR</td>
<td>Most of the CFQoL variables were moderately correlated (r=0.38–0.61, p&lt;0.001) with the single item scale weakly correlated with body image (r=0.25), p&lt;0.01 Higher scores correlated negatively with frequency of hospital admissions in the previous year (r=−0.39, p&lt;0.001)</td>
<td>Ability to distinguish adult CF patients with lower compared to higher CFQoL scores</td>
<td>Single-item scale correlation with FEV1 r=0.21</td>
<td>NR</td>
<td>ICC 0.78 (95% CI 0.59–0.88)</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>Acceptable, valid and repeatable measurement tool that can be easily used</td>
</tr>
<tr>
<td>CQOLCF</td>
<td>NR</td>
<td>Correlation with mental health r=0.634, emotional distress r=−0.687 and physical health r=0.049</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>Cronbach’s α=0.909</td>
<td>NR</td>
<td>NR</td>
<td>Appears to be valid, reliable and internally consistent scale</td>
</tr>
</tbody>
</table>

MCID: minimal clinically important difference; CFIQ: CF Impact Questionnaire; CFQoL: CF QoL Questionnaire; CFQ-R: revised CF Questionnaire; CF-QUEST: electronic version of the CFQoL evaluative self-administered test; FLZ-CF: Questions of Life Satisfaction; CQOLCF: Caregiver QoL CF scale; NR: not reported; SF-36: Short-Form-36 Item Questionnaire; HRQoL: health-related QoL; CFTR: CF transmembrane regulator; FEV1: forced expiratory volume in 1 s; GHQ: General Health Questionnaire; KINDL-R: Child QoL Questionnaire-Revised; ICC: internal consistency coefficient; rs: Spearman’s correlation coefficient.
<table>
<thead>
<tr>
<th>Test or tool</th>
<th>Content validity</th>
<th>Convergent validity</th>
<th>Discriminant validity</th>
<th>Concurrent validity</th>
<th>Predictive validity</th>
<th>Intra or inter-rater test-retest</th>
<th>Internal consistency</th>
<th>Measurement error</th>
<th>Responsiveness</th>
<th>Comments/MCID</th>
</tr>
</thead>
</table>
| **Pulmonary exacerbations**

<table>
<thead>
<tr>
<th>Test or tool</th>
<th>Content validity</th>
<th>Convergent validity</th>
<th>Discriminant validity</th>
<th>Concurrent validity</th>
<th>Predictive validity</th>
<th>Intra or inter-rater test-retest</th>
<th>Internal consistency</th>
<th>Measurement error</th>
<th>Responsiveness</th>
<th>Comments/MCID</th>
</tr>
</thead>
<tbody>
<tr>
<td>AWESCORE</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>Correlation of total AWESCORE and CFQ-R scores: r=0.632 (p=0.001)</td>
<td>NR</td>
<td>Pearson’s correlation coefficient 0.854, p&lt;0.0005</td>
<td>NR</td>
<td>NR</td>
<td>For exacerbation, score 47.5 (SD 11.2) at start of treatment and 21.6 (SD 15.6) at end of treatment (100=highest symptom severity)</td>
<td>11 points Mean change of −16.5 (95% CI −13.2 to −19.7 for exacerbation reported) No MCID for emotional score</td>
</tr>
<tr>
<td>CFRSD/CFRSD-CRISS</td>
<td>Involved people with CF in testing clarity of items</td>
<td>Step-rate significantly higher in those who did NOT experience difficulty breathing, cough, tightness or feeling tired (respiratory items) or feeling worried, cranky or frustrated (emotional items)</td>
<td>Respiratory scores distinguished between moderate/severe and mild/severe disease; emotional scores distinguished between mild/severe disease</td>
<td>Respiratory and emotional score established using daily step count (not FEV₁)</td>
<td>ICC 0.79 for respiratory scale using a 1-day interval</td>
<td>Cronbach’s α for CFRSD-CRISS was 0.77</td>
<td>Test-retest reliability after 7–10 days; 0.9 for the emotional and 0.93 for the chest score</td>
<td>Total score been demonstrated to improve over 2 weeks’ i.v. treatment</td>
<td>No MCID suggested on the basis of statistical analysis, but MCID &gt;1 after 2 weeks of i.v. ABX suggested based on experience with COPD patients</td>
<td></td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>Test or tool</th>
<th>Content validity</th>
<th>Convergent validity</th>
<th>Discriminant validity</th>
<th>Concurrent validity</th>
<th>Predictive validity</th>
<th>Intra or inter-rater and test-retest</th>
<th>Internal consistency</th>
<th>Measurement error</th>
<th>Responsiveness</th>
<th>Comments/ MCID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symptom score</td>
<td>Patients not involved in construction</td>
<td>All 4 items correlated with each other ( r=0.38; \ p&lt;0.001 )</td>
<td>NR</td>
<td>Total score correlation with FEV(_1): ( r=-0.41; \ (p&lt;0.0001) ) and respiratory score on CFQ-R: ( r=-0.62; \ (p&lt;0.001) ) and CFQ-R: ( r=-0.47; \ (p&lt;0.001) )</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
</tbody>
</table>

**Respiratory**

<table>
<thead>
<tr>
<th>Borg Dyspnoea Scale</th>
<th>NR</th>
<th>NR</th>
<th>NR</th>
<th>NR</th>
<th>NR</th>
<th>ICC=0.933</th>
<th>NR</th>
<th>NR</th>
<th>Mean change in score (-3.1) with mean effect size (1.3) from baseline to 4 weeks</th>
<th>Appears to be valid, reliable and responsive in CF For those reporting improvement, scores changed (-2.9) overall, (-3.5) for cough, (-3.5) for congestion and (-3.1) for sputum domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>ReS-CF</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>Correlation between ReS-CF and CFQ-R: ( r_c=-0.5; \ (p&lt;0.001) )</td>
<td>NR</td>
<td>ICCs for 4 scores &gt;0.7</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
</tbody>
</table>

---

https://doi.org/10.1183/16000617.0354-2020
<table>
<thead>
<tr>
<th>Test or tool</th>
<th>Content validity</th>
<th>Convergent validity</th>
<th>Discriminant validity</th>
<th>Concurrent validity</th>
<th>Predictive validity</th>
<th>Intra or inter-rater and test-retest</th>
<th>Internal consistency</th>
<th>Measurement error</th>
<th>Responsiveness</th>
<th>Comments/MCID</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOBQ</td>
<td>NR</td>
<td>SOBQ scores correlated negatively with physiological measures of disease severity (FVC % pred: r=−0.36, p&lt;0.05 and FEV₁ % pred: r=−0.5, p&lt;0.01)</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>α=0.96</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>MCID: 5 unit change</td>
</tr>
</tbody>
</table>

SOBQ scores correlated positively with Borg scale ratings of perceived breathlessness after 6MWT and QWB (r=−0.41, p=0.01)
<table>
<thead>
<tr>
<th>Test or tool</th>
<th>Content validity</th>
<th>Convergent validity</th>
<th>Discriminant validity</th>
<th>Concurrent</th>
<th>Predictive</th>
<th>Intra or inter-rater and test-retest</th>
<th>Internal consistency</th>
<th>Measurement error</th>
<th>Responsiveness</th>
<th>Comments/MCID</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pain</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BPI</td>
<td>NR</td>
<td>Correlation of BPI pain interference and airway clearance therapy (p=0.002), coughing and breathing (p=0.012), pain prevalence and CFQoL physical function (p=0.01), CFQoL treatment (p=0.03), CFQoL work/school (p=0.02), CFQoL social (p=0.013) and CFQoL emotional scale (p=0.017) Pain intensity also correlated with CFQoL physical function, CFQoL treatment and CFQoL school/work (p&lt;0.01)</td>
<td>NR</td>
<td>Correlation of BPI pain prevalence and sleep quality (p=0.045), sleep disturbance (p=0.001), daytime dysfunction (p=0.001) and sleep interference and global BPI score rho-0.56, p&lt;0.0001 OR 1.27 (p=0.012) of impaired sleep quality in those with pain</td>
<td>BPI pain severity correlated with risk of exacerbations (OR 1.65, p=0.04) for exacerbations with higher pain intensity and OR of 2.28 (p=0.008) of death with higher pain intensity</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Test or tool</td>
<td>Content validity</td>
<td>Convergent validity</td>
<td>Discriminant validity</td>
<td>Concurrent</td>
<td>Predictive</td>
<td>Intra or inter-rater and test-retest</td>
<td>Internal consistency</td>
<td>Measurement error</td>
<td>Responsiveness</td>
<td>Comments/ MCID</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------</td>
<td>---------------------</td>
<td>-----------------------</td>
<td>------------</td>
<td>-----------</td>
<td>---------------------------------</td>
<td>---------------------</td>
<td>-----------------</td>
<td>---------------</td>
<td>----------------</td>
</tr>
<tr>
<td>DPAQ-CF</td>
<td>NR</td>
<td>CFQ-R social function (r=0.269, p&lt;0.01), CFQ-R treatment burden (r=0.269, p&lt;0.01), CFQ-R respiratory symptoms (r=0.241, p&lt;0.05), HADS-depression scale (r=0.29, p&lt;0.01) and HADS-anxiety (r=0.29, p&lt;0.01)</td>
<td>DPAQ-CF pain intensity correlated with CFQ-treatment burden and respiratory symptoms (p&lt;0.01)</td>
<td>DPAQ-CF pain duration correlated with CFQ-R treatment burden and respiratory symptoms (p&lt;0.01)</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
</tbody>
</table>

NR Correlation of pain and ppFEV₁ R=0.239, P<0.05

Continued
<table>
<thead>
<tr>
<th>Test or tool</th>
<th>Content validity</th>
<th>Convergent validity</th>
<th>Discriminant validity</th>
<th>Concurrent</th>
<th>Predictive</th>
<th>Intra or inter-rater and test-retest</th>
<th>Internal consistency</th>
<th>Measurement error</th>
<th>Responsiveness</th>
<th>Comments/ MCID</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPI</td>
<td>NR</td>
<td>Correlation of BPI pain severity and Shwachman scale history scale; ( r=0.24 ) (( p=0.04 )) and BPI pain interference and total Shwachman score ( r=0.2 ) (( p=0.09 ))</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abdominal

| CF-Abd score | NR | NR | Differentiated patients with CF and healthy controls with large effect size (17.3±1.1 versus 8.0 ±0.7 points; \( p<0.001 \); Cohen’s d=0.85) | NR | NR | ICC 0.932 (95% CI 0.874–0.963) | Cronbach’s \( \alpha=0.7–0.92 \) | NR | NR |

Gastrointestinal symptom tracker

| NR | NR | Nutritional status is related to more stable lung function and fewer exacerbations | NR | NR | NR | Reliability established based on test–retest and internal consistency (unspecified) | NR | NR | NR |

Continued
### TABLE 5 Continued

<table>
<thead>
<tr>
<th>Test or tool</th>
<th>Content validity</th>
<th>Convergent validity</th>
<th>Discriminant validity</th>
<th>Concurrent Predictive</th>
<th>Intra or inter-rater and test-retest</th>
<th>Internal consistency</th>
<th>Measurement error</th>
<th>Responsiveness</th>
<th>Comments/MCID</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Symptom and impact score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSAS-CF</td>
<td>Developed in accordance with COSMIN recommendations; patients not consulted</td>
<td>MSAS-Resp: correlation with CFQ-R (r = −0.60, p &lt; 0.05) and CFQoL (r = −0.7, p &lt; 0.05)</td>
<td>Respiratory, gastrointestinal and psychiatric scores were higher in patients with low FEV₁ &lt;40% pred (p &lt; 0.05)</td>
<td>Correlation with CFQ-R respiratory score (r = −0.6) and CFQoL chest score (r = −0.7, p &lt; 0.05) and CFQ-R emotional functioning score (r = −0.69, p &lt; 0.05).</td>
<td>NR</td>
<td>NR</td>
<td>α 0.74–0.86</td>
<td>High in all domains; MSAS-Physical α 0.92, MSAS-Psych α 0.95, MSAS-Global α 0.82</td>
<td>NR</td>
</tr>
</tbody>
</table>

MCID: minimal clinically important difference; AWESCORE: Alfred Wellness Score; CFRSD: CF respiratory symptom diary; CRISS: chronic respiratory infection symptom score; ReS-CF: respiratory symptoms in CF tool; SOBQ: Shortness of Breath Questionnaire; BPI: brief pain inventory; DPAQ-CF: Daily Pain Assessment Questionnaire in CF; MPI: multiple pain inventory; CF-Abd: CF abdominal; MSAS: Memorial Symptom Assessment Scale; NR: not reported; FEV₁: forced expiratory volume in 1 s; FVC: forced vital capacity; ICC: internal consistency coefficient; ABX: antibiotics; 6MWT: 6-min walk test; QWB: Quality of Well-Being Questionnaire; HADS: Hospital Anxiety and Depression Scale; Resp: respiratory; GI: gastrointestinal; CFQ-R: CF Questionnaire-revised; COSMIN: Consensus-based Standards for the selection of health Measurement Instruments initiative.
discriminate between people with CF and healthy controls, as well as those at different disease stages based on age, infection and structural abnormalities identified on high-resolution CT imaging or MRI [110]. A correlation with clinical outcomes has not been established.

**Imaging scoring tools**

While the Brasfield and Wisconsin CXR scores performed similarly and both have been found to be reproducible (intra-observer agreement r=0.86–0.99 and 0.78–0.96, respectively) and reliable (inter-rater agreement 0.76–0.90 and 0.74–0.97, respectively), they appear to be insensitive to early disease [122]. The correlation between these scores was reported as r=0.86, p<0.0001. Both scores correlated with lung function (FEV₁ and forced vital capacity, all p<0.001) [73]. The correlation of scores with FEV₁ was highest for the Northern score (r=−0.82) compared to the Brasfield (r=0.81) or Crispin–Norman scoring methods (r=−0.83) [124].

CT scoring tools have been found to have higher sensitivity for detecting lung disease progression than FEV₁ % pred. The test–retest reliability based on the intraclass correlation coefficient of the PRAGMA-CF score was shown to be >0.9 for percentage disease, 0.85 for percentage bronchiectasis and 0.96 for percentage air trapping; the intra-observer reliability was >0.90 for bronchiectasis, air trapping and percentage disease [50].

The test–retest reliability of a semi-quantitative MRI score was $r^2=0.76$ (p=0.0047) [54] and correlation with FEV₁ was $r=0.81$ (p=0.0023) [54].

**Functional exercise performance**

A summary of the measurement properties of tests used to capture functional exercise capacity is provided in table S6.

Many tests capturing functional exercise performance were compared to cardiopulmonary exercise testing (CPET), which has historically been viewed as the gold standard for assessing exercise capacity according to Von Berg et al. [57]. Rand et al. [47] found that the incremental field step test had acceptable concurrent validity compared to CPET in children for measuring peak oxygen uptake, minute ventilation, heart rate, change in oxygen saturation and CO₂ ventilation and perceived exertion [47].

Submaximal exercise tests included the 6MWT, 3-min step test (3MST) and modified shuttle walk test (MSWT) and 30-s or 1-min sit-to-stand test [57]. Good concurrent validity of the MSWT with maximum oxygen capacity on CPET has been reported; however, results for concurrent validity were inconsistent for the 6MWT and 3MST. The ability of the 6MWT to predict pre-transplant survival was variable [36]. A reduction of 50 m or more in the modified shuttle test was associated with a hazard ratio of death or lung transplant within 1 year in adults with CF of 1.91 (95% CI 1.09–3.35, p<0.024) [20]. Convergent validity of 3MST and MSWT with FEV₁ (r=0.61, p=0.002) was found [36], but this was variable for the 6MWT.

**CFTR function tests**

Intestinal current measurement and nasal potential difference (NPD) tests, which directly measure CFTR function, were strongly correlated and have been found to distinguish people with CF from healthy controls (k=0.83 versus k=0.33, respectively, p<0.001) [68]. Changes in NPD have been reported over 14 days in trials of the CFTR function-modifying drug ivacaftor. Some evidence for the reliability of intestinal organoid volume has been found, but evidence to support its validity has not [68]. Some evidence for the validity and reliability of indirect measures of gastrointestinal CFTR function such as intestinal pH, faecal calprotectin and faecal elastase-1 has been found; however, these data are not described in detail in the review included in our study (table S6).

**Sputum tests**

Tests characterising sputum rheology, including viscoelasticity and solid content properties, demonstrated poor to fair test–retest reliability with ICCs ranging from 0.22 to 0.42 (with wide confidence intervals) [46, 103]. Reproducibility of biomarkers in the sputum such as total cell count, neutrophils, tumour necrosis factor-α, interleukin-8 and neutrophil elastase was demonstrated in one study [128] as follows: ICC=0.76, 0.82, 0.93, 0.82 and 0.74, respectively; however, there was marked between-patient variability [103, 128].

**Measurement error**

The systematic and random error of a patient’s score not attributable to true changes in the construct that was measured was poorly reported across all studies (table 4, table 5 and table S6).
Discussion

While the measurement properties of PROMs evaluating HRQoL in CF studies have been previously evaluated [135], this is the first effort to systematically review evidence of the measurement properties of all tests and tools used in CF studies. A diverse range of tests and tools were identified which vary with respect to their reliability, responsiveness and validity. There was inconsistency in the use of tests and tools to measure the same or similar outcomes across studies. This highlights the need to establish consensus over which outcomes should be measured in CF studies and how they should be measured; this has been recommended by the COSMIN initiative group [136]. Compared to older tools, many recently developed tools incorporate self-reported outcomes by patients (e.g. CFRSD-CRISS, CFQ-R, CF Impact Questionnaire and CF Quality of Life (CFQoL)) and have involved people with CF in their development, consistent with the recommendation made by the US FDA in 2017 [137].

Evidence to support the reliability of spirometry testing was found; this has also been substantiated in other populations, such as in people with other chronic lung disease [138]. Poor FEV₁ is strongly correlated with death, progression to lung transplant (most transplant recipients have a FEV₁ <30% pred) [139] and reduced QoL [110] in people with CF and is also associated with a greater risk of hospitalisation, pulmonary exacerbations and colonisation with Pseudomonas aeruginosa [140]. Compared to crude or percentage predicted FEV₁ values, z-scores have been proposed as a less biased and more accurate measure for defining meaningful changes in lung function since they take into account sex and ethnicity in addition to age and height; this approach has been endorsed by the Global Lung Initiative since 2013 [141]. This, however, has not yet been universally adopted as the preferred measure for capturing lung function in CF studies. Consensus regarding the MCID for FEV₁ was not identified in this review, but MCIDs have been proposed. In the TRAFFIC and TRANSPORT phase 3 trials, which evaluated lumacaftor–ivacaftor versus placebo for people homozygous for the Phe508del CFTR mutation [142, 143], a mean relative difference of 3.3% (2.3–4.4, p<0.0001) and 2.8% (1.7–3.8, p<0.0001) was found in those with baseline FEV₁ ≥40% pred and baseline FEV₁ <40%, respectively. It was proposed that this represents a clinically significant improvement since the annual rate of decline of FEV₁ % pred has been estimated to be 1.92% per annum for people with CF aged 1824 years (n=2793) and 1.45% for those aged >25 years [144].

While FEV₁ has been shown to be reproducible and repeatable in children aged ≥6 years and adults, its variability is affected by the person’s age and the severity of their underlying lung disease [110]. In the early stages of CF disease, FEV₁ often remains within the normal range, while in severe lung disease FEV₁ is significantly compromised and unlikely to demonstrate variability [89]. LCI testing represents an alternative test for children aged <6 years who are incapable of performing spirometry. Since measurement is dependent on body size, the relative rather than the absolute change is considered more appropriate, at least before 6 years of age [145]. LCI has been shown to correlate strongly with structural abnormalities detected on high-resolution CT and abnormal preschool LCI is associated with spirometry deficits performed within 3 years from baseline in school-age children [146]. However, further standardisation and evaluation of the relationship of LCI with morbidity and mortality is warranted.

Evidence of the reliability, responsiveness and validity of two commonly used QoL tools, the CFQ-R and the CFQoL, as well as the CFRSD-CRISS symptom scoring tool has been reported previously and has been substantiated by this review. The content validity (including face validity) of these tools is consistent with the recommendation made by the US FDA in 2017 [137]. There have been significant advances in treatment and long-term health outcomes for people with CF in recent decades, which raises a concern about the current content validity of some of the outcome scoring tools developed in the second half of the 20th century, many of which did not involve people with CF in their development [43]. Many of these have not undergone sufficient validation and consequently have not been recommended for use in clinical practice or in research.

The use of imaging modalities and scoring tools in CF has evolved with time; however, considerable variability exists between treatment centres, for example whether to use CXR or CT for longitudinal disease monitoring. An important limitation of CXR imaging is its poor sensitivity for detecting structural lung changes in early disease and progression in those with established disease [73]. This modality, however, is still used for monitoring disease progression in some treatment centres, and it has an established role in identifying pathology in the context of an acute clinical deterioration, such as consolidation or pneumothorax. Extensive collaboration has occurred within the CF community to standardise CT and MRI radiological scores, especially in young children, to enable quantification of the degree of structural lung damage. While CT is currently the most sensitive method for detecting structural
airways disease [147]. MRI shows promise because it delivers non-ionising radiation and allows assessment of functional aspects of the lung such as perfusion, pulmonary haemodynamics and ventilation [111]. It may be possible to automate imaging scoring algorithms in the future, which may improve the efficiency and reliability of results. However, further assessment of the validity and reproducibility of MRI scoring tools is required, and the extent to which imaging scores predict clinical outcomes of significance requires further elucidation, including in children [111].

The strengths of this review include the use of a systematic approach to identify studies by two independent reviewers. There were four major limitations. First, tests and tools used in practice in people with CF that have been validated in non-CF populations (e.g. generic scores capturing abdominal symptoms) were considered beyond the scope of this review. Secondly, details about the systematic error (bias) and random error (noise) for each of the tests and tools (i.e. variation beyond that attributable to the outcome of interest) have been poorly described in the literature. Measurement error is an important source of bias; this information is necessary to appraise the quality of tests and tools and should be an important factor influencing selection. Thirdly, medical devices used to capture outcomes were beyond the scope of this review (such as weighing scales or stadiometers used to capture anthropometric outcomes). Finally, given the large scope of this review, an exhaustive critique of the measurement properties of individual tests and tools was not feasible.

Conclusions

This systematic review highlights the diversity of tests and tools which have been used for outcome measurement in CF studies and their variable characteristics and properties. While there have been concerted efforts within the CF research community to improve and standardise these tests and tools, further work is needed, particularly to optimise tools for outcome measurement in young children and those with mild or severe disease. A consensus set of tests and tools for measurement in CF studies is needed; this should be developed together with people with CF and other relevant stakeholders. This would likely improve the consistency of reporting and measurement of similar outcomes, allowing comparison and synthesis of evidence across studies and improving the value of the research that is conducted.

Provenance: Submitted article, peer reviewed

Acknowledgements: This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors. C. McLeod is supported by an NHMRC post-graduate scholarship (GNT1150996), and top-up grants from the Wesfarmers Centre and the Perth Children’s Hospital Foundation (9772 and 9757). T.L. Snelling is supported by a Career Development Fellowship from the National Health and Medical Research Council (GNT1111657). C.C. Blyth is also supported by the National Health and Medical Research Council (1111596/1173163).

Author contributions: C. McLeod was responsible for the study conceptualisation and overall methodology. C. McLeod and J. Wood were responsible for data curation. T.L. Snelling, C. McLeod, S. Smith, S. Webb, C.C. Blyth, A. Tong, J. Wood, A. Schultz and A.R. Smyth elaborated the study protocol. C. McLeod drafted the manuscript. All authors were involved in the interpretation of data and revision of the manuscript. All authors approved the final manuscript.

Conflict of interest: C. McLeod has nothing to disclose. J. Wood has nothing to disclose. A. Tong has nothing to disclose. A. Schultz reports personal fees from Vertex Pharmaceuticals, outside the submitted work. R. Norman has nothing to disclose. S. Smith has nothing to disclose. C.C. Blyth has nothing to disclose. S. Webb has nothing to disclose. A.R. Smyth reports grants from Vertex, speaker honoraria and expenses from TEVA and Novartis, and personal fees from Vertex, outside the submitted work. In addition, A.R. Smyth has a patent “Alkyl quinolones as biomarkers of Pseudomonas aeruginosa infection and uses thereof” issued. T.L. Snelling has nothing to disclose.

Support statement: Funding was received from Perth Children’s Hospital Foundation (grants 9772 and 9757), National Health and Medical Research Council (GNT1111657) and Wesfarmers Centre for Vaccines and Infectious Diseases. Funding information for this article has been deposited with the Crossref Funder Registry.

References


Daily pain in adolescents with CF: effects on adherence, psychological symptoms, and health-related quality of life. 


U.S. Food and Drug Administration. Table of surrogate endpoints that were the basis of drug approval or-licensure. 

American Thoracic Society. Cystic Fibrosis Questionnaire (CFQ) and Cystic Fibrosis Questionnaire Revised (CFQ-R). 

Ratnayake I, Ahern S, Ruseckaite R. A systematic review of patient-reported outcome measures (PROMs) in cystic fibrosis. 

Cystic Fibrosis Foundation Patient Registry. 2018 Annual Data Report. 


COSMIN. www.cosmin.nl/ 


Wisconsin cystic fibrosis chest radiograph scoring system. 

Cystic Fibrosis and lung transplantation – determination of the survival benefit. 

The chest radiograph in cystic fibrosis: a new scoring system compared with the Chrispin-Norman and Brasfield scores. 

The Global Lung Function Initiative: dispelling some myths of lung function test interpretation. 


The chest radiograph in cystic fibrosis: a new scoring system. 

Cystic Fibrosis Questionnaire (CFQ) for assessing quality of life in pediatric and adult patients. 

Prediction of mortality in patients with cystic fibrosis. 

Published: 2020-09-30


Chapter 6: Protocol for a discrete choice experiment (DCE)

6.1 Chapter summary

This chapter presents a protocol for the design, conduct and analysis of a discrete choice experiment (DCE) for quantifying how people with CF value different health states, comprising a range of attributes (outcomes) resulting from treatment of pulmonary exacerbation episodes.

Attributes for inclusion in the DCE were chosen based on the top ten attributes that captured the impact on symptoms or function that (i) were identified as meaningful to those with lived experience of the disease (ii) assessed separate rather than shared aspects of the underlying pathophysiological disease processes and (iii) were likely to be impacted by treatment interventions for pulmonary exacerbations of CF.

An objective of the DCE was to determine the weights, or relative importance of the ten prioritized attributes for inclusion in a MAOI; one instrument for summarizing proxy carer-reported outcomes measures for use in children with CF and another instrument for summarising patient-reported outcome measures for use in people older than 13 years with CF.

6.2 Journal article


Appendix 1 is available electronically here.
BMJ Open

Discrete choice experiment to evaluate preferences of patients with cystic fibrosis among alternative treatment-related health outcomes: a protocol

Charlie McLeod, Richard Norman, Andre Schultz, Steven Mascaro, Steve Webb, Tom Snelling

ABSTRACT

Introduction Clinical decision-making is a complex process. Patient preference information regarding desirable health states should inform treatment and is critical to agreeing on goals of therapy. Cystic fibrosis (CF) is a common, inheritable multisystem disorder for which the major manifestation is progressive, chronic lung disease. Intermittent pulmonary exacerbations are a hallmark of disease and these drive lung damage that results in premature death. We suspect that clinicians make assumptions, most likely implicit assumptions, about outcomes that are desired by patients who are treated for pulmonary exacerbations. The aim of this study is to identify and quantify the preferences of patients with cystic fibrosis regarding treatment outcomes.

Methods and analysis We will develop a discrete choice experiment (DCE) in collaboration with people with CF and their carers, and evaluate how patients make trade-offs between different aspects of health-related status when considering treatment options.

Ethics and dissemination Ethics approval for all aspects of this study was granted by the Western Australia Child and Adolescent Health Service Human Research Ethics Committee [RSS903]. Weighted preference information from the DCE will be used to develop a multiattribute utility instrument as a measure of treatment success in the upcoming Bayesian Evidence-Adaptive Trial to optimise management of CF. Dissemination of results will also occur through peer-reviewed publications and presentations to relevant stakeholders and research networks.

INTRODUCTION

Medical decision-making is a complex process. In the clinical setting, this should be a shared, iterative process between clinicians and patients (and their carers if appropriate). Each group brings differing needs and perspectives. Understanding patient preferences regarding health outcomes is critical to informing treatment choices and agreeing to goals of therapy. In addition to being desired by patients, these goals must also be considered achievable by clinicians.

Cystic fibrosis (CF) occurs in 1:2000 to 1:3500 births and is an inheritable multisystem disorder for which the major manifestation is progressive, chronic lung disease. Survival improved dramatically during the latter part of the 20th century but has more recently slowed with average survival approximately 50 years. The disease is characterised by intermittent pulmonary exacerbations which drive lung damage. Minimising the decline in lung function that accompanies pulmonary exacerbations (one in four patients do not recover their baseline function) is thought to be key to improving survival and quality of life. Management of pulmonary exacerbations generally involves a combination of antimicrobial, anti-inflammatory and mucolytic agents, physiotherapy and optimisation of nutrition. However, there is no consensus between centres regarding a standardised approach due to the paucity of evidence available to guide therapy.
The James Lind Alliance, in partnership with healthcare providers and people with CF from 23 countries, has recognised treatment of pulmonary exacerbations as a research priority and specific knowledge gaps in this area have been recently identified.

Determining the value of different treatment options depends on the value patients place on the consequences of each treatment decision. A variety of methods exist to elicit patient preference information. These include revealed preference and stated preference methods. Revealed preferences are based on observed choices made by individuals in real-life scenarios. Stated preferences are derived from decisions made by individuals when confronted with realistic, hypothetical choice scenarios, such as in a discrete choice experiment (DCE). DCEs can also capture drug benefit vs toxicity, compared with other techniques where older participants are present, workshops will be conducted separately for the following groups: young individuals, adults with CF (13–25 years), adults with CF (>25 years) and administration of DCE to weigh the relative importance of outcomes from the perspective of patients. This will allow the relative weights of evaluated outcomes to be determined.

DCE involves administration of a choice-based questionnaire that presents clinical vignettes and asks respondents to make trade-offs between different aspects of health-related status. The core theory informing DCE design is that the value of an option depends on the value of its attributes. Attributes are characteristics of treatments or services that may be processes (factors related to the delivery of care), structures (such as the setting in which healthcare occurs) and/or health outcomes, which are considered most important to people living with disease. Weighted patient preference information from the DCE will be incorporated into a multiattribute utility instrument (MAUI), which will generate a score as a measure of success in pulmonary exacerbation trials, including the planned Bayesian Evidence-Adaptive Trial to optimise management of CF (BEAT-CF). We expect recruitment for this study will largely occur within Australia, which may limit the generalisability of findings to other CF populations.

The aims of this study are (1) to identify and prioritise health outcomes of importance to people affected by CF, (2) to map these outcomes to consensus-derived causal models of CF pulmonary exacerbations and (3) to examine how patients make trade-offs between different aspects of health-related status when considering treatment decisions.

**METHODS AND ANALYSIS**

**Overview of approach and consumer involvement**

Consumer involvement is critical to this work which will comprise four stages (figure 1): (1) key health outcome elicitation and prioritisation from the perspective of people affected by CF, (2) mapping of these outcomes to a consensus-derived causal model of the disease processes, (3) selection of outcomes for inclusion in the DCE taking into consideration orthogonality and (4) development and administration of DCE to weigh the relative importance of outcomes from the perspective of patients. This study commenced in October 2018 and completion is expected in May 2020. Study progression at each stage is contingent on completion of the preceding research stage.

**Key health outcome elicitation by CF consumers**

Elicitation of key health outcomes from consumers will occur using two methods: (1) preliminary consumer workshops and (2) online health outcomes surveys. Patient preference information is expected to vary between individuals but also according to age and stage of disease. To help elucidate these differences, and because young people may be less inclined to contribute in a group where older participants are present, workshops will be conducted separately for the following groups: young people with CF (13–25 years), adults with CF (>25 years)
and persons who identify as carers for people with CF (including parents).

Workshops for patients will occur via teleconference, owing to infection control restrictions which preclude direct contact among this patient population. Carer workshops will be conducted in-person at the Telethon Kids Institute (Perth, Australia) with teleconference dial-in facilities available if requested.

Follow-up workshops
Outcomes identified through the consumer engagement activities detailed above will be collated with any additional potentially important health outcomes identified from review of the literature. Prioritisation of outcomes will occur during a series of follow-up workshops with each of the consumer groups. A combined workshop will also be conducted to derive a consensus list of prioritised outcomes relating to treatment of pulmonary exacerbations from the perspective of patients >13 years.

Consensus causal diagram
A consensus causal model (in the form of a Bayesian network) which links outcomes to causal disease processes for pulmonary exacerbations will be developed by a group of clinicians and other subject domain experts and people with lived experience of the disease. This process will be moderated by external facilitators using expert knowledge elicitation methods. The purpose of this is to guide selection of outcomes for inclusion in the DCE by choosing those that are likely to be important while minimising the inclusion of multiple attributes that measure the same outcome. The causal model will also aid in identifying probable combinations of attributes to ensure they are covered by the DCE, as well as helping to rule out improbable attribute combinations. Finally, the model will identify dependencies between attributes that need to be controlled for or otherwise handled during the analysis.

DCE design
The first step in designing the DCE is the identification of the important attributes (characteristics) for evaluation, and the assignment of possible levels to these attributes.

Attributes and levels
Attributes and levels will be selected according to guidance provided by ISPOR. Only attributes that are identified as important to people with CF that map to causal disease pathways will be considered for inclusion. Levels (which may be categorical, continuous or probabilities) will be assigned in consultation with consumer representatives based on those that patients can relate to and consider meaningful which best represent the spectrum of possibilities that are clinically encountered.

Code to generate design
An experimental design will be constructed chiefly by RN in Ngene, software widely used in DCE development. The principles underpinning our design is that it will (1) consist of a pool of choice tasks, divided into blocks to which respondents will be randomly allocated, (2) maximise efficiency in terms of the precision of the coefficients (ie, D-efficiency) and (3) account for the ordered nature of the parameters under consideration by employing small non-zero priors in Ngene. As described below, the design may be updated following qualitative review of the initial design.

DCE questionnaire
The questionnaire will contain background information explaining the study rationale and potential risks and benefits of participating. Attributes and levels will be clearly defined. Sociodemographic data (age, sex, postcode) will be collected to assess if these factors influence stated preferences.

The draft DCE will be administered to a convenience sample of consumers (see figure 2 for DCE choice task example). If the tasks are too difficult or present implausible combinations of levels, we will define a candidate set of acceptable choice sets, and regenerate the design with a fixed amount of level overlap. Feedback about other design elements, including the length, layout, specific wording and comprehensibility will also be obtained. Suggestions for improvement will be considered and the final model agreed by consensus.

Sampling and recruitment strategy
There are 3,422 people registered on the National CF database in Australia. Our research population comprises patients ≥13 years with CF and individuals who identify as carers for person(s) living with CF.

Recruitment from the sampling pool for stages 1 and 3 of this study will occur through a variety of means including through outpatient clinics and inpatient wards at Sir Charles Gairdner Hospital (adult tertiary hospital facility) and Perth Children’s Hospital (children’s tertiary hospital facility) and by advertising through consumer and research networks, including the Western Australia CF consumer reference group, CF Australia and CF Western Australia, including through electronic and social media bulletins and communiqués. Interested persons will contact a member of the study team by phone or email to register their interest. Patient information and consent forms will be provided for the workshop (electronically via email attachment or wet signature for in-person workshop attendees) and survey participants (online). Participants aged between 13 and 18 years will additionally require guardian consent. Links for the online CF-related health outcomes survey and DCE questionnaire will be sent via email once consent forms are received.

Workshops will proceed if two or more consumers register to attend. For the combined workshop, we aim to recruit a minimum of two young people and two adults with CF.

There is no consensus regarding DCE sample size requirements for applications in healthcare. ISPOR guidance remarks that statistical precision increases at sample size.
Please consider the following two options for treatment of a CF exacerbation versus the option to have no treatment. The different treatment options could have a different effect on your lung function and how you are feeling, and there may be side-effects resulting from treatment.

<table>
<thead>
<tr>
<th></th>
<th>TREATMENT A</th>
<th>TREATMENT B</th>
<th>NO TREATMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change in lung function</td>
<td>+10% (good improvement)</td>
<td>+5% (a little better)</td>
<td>0% (no change)</td>
</tr>
<tr>
<td>How you feel</td>
<td>Excellent</td>
<td>Fair</td>
<td>Fair</td>
</tr>
<tr>
<td>Side-effects of antibiotic treatment</td>
<td>Moderate increase in dry cough during, and for up to 15 minutes after each nebuliser</td>
<td>Stomach cramps and watery diarrhoea up to 4 times per day</td>
<td>None</td>
</tr>
<tr>
<td>Effect on school/work</td>
<td>Able to return to school/work at 50% capacity after 1 week, then 75% for 4 weeks, then 100%</td>
<td>Unable to return to school/work for 2 weeks, the return to school/work at 50% capacity for 2 weeks, then 100%</td>
<td>Unable to return to school/work for 3 weeks, the return to school/work at 50% capacity for 2 weeks, then 100%</td>
</tr>
</tbody>
</table>

![Figure 2](https://doi.org/10.1136/bmjopen-2019-030348) Discrete choice experiment choice task example. CF, cystic fibrosis

sizes above 150 and levels out over 300 observations.\(^{23}\) Lancsar et al suggests a minimum of 20 observations per choice set is required to achieve a reliable model,\(^ {18}\) while Marshall provides a rough rule of thumb based on the number of tasks, alternatives per choice set and levels.\(^ {34}\) DCE will remain open until 200 responses are received and 4 months have elapsed since commencement. This target sample size represents a compromise between the desire for an accurate tool (one that reflects the average preferences for consumers) and the practical consideration that, at most, we aspire for roughly 1 in 10 patients to contribute from the sample pool of approximately 2000 people >13 years with CF.\(^ {33}\)

Given our recruitment strategy, participants are expected to predominantly reside in Australia, although it is possible that some participants living overseas may participate, depending on the reach of our consumer and research networks. As a robustness check, analyses will be conducted with and without any non-Australia-based respondents.

### Participant reimbursement

Participants will not be paid to take part in any aspect of this study. Parking reimbursement for those who attend the in-person caregiver workshops will be provided.

### Patient and public involvement

BEAT-CF will focus on evaluating optimal treatment(s) for pulmonary exacerbations, which has been identified by the James Lind Alliance as a research priority for people affected by CF.\(^ {13}\) Consumer advocates have been involved in elements of trial design, and patients will be involved at all stages of the research process. Patients are not officially involved in participant recruitment, although promotion of research activities is expected to occur by word-of-mouth. Results will be disseminated to participants involved in this study and broadly via peer-reviewed presentations and by consumer research networks and CF advocacy organisations.

### Data collection

Workshops will be approximately 2 hours in duration. Outcome elicitation (preliminary workshops) will occur using nominal group technique.\(^ {35}\) Key aspects of this approach include clarification of the purpose of the session, allowing time for participants to formulate individual responses, and then asking participants to present one idea aloud, in turn to the group until saturation occurs, that is, until no new outcomes are identified.\(^ {35}\) Results for these sessions will be collated on Excel spreadsheets and remain visible to participants throughout the session. Discussion
of individual ideas will be permitted to allow clarification, rather than to resolve differences. A facilitator will ensure discussion is equally balanced among all ideas and between individuals. Prioritisation of outcomes (follow-up workshops) will occur by collating results from participants asked to rank outcomes at the follow-up workshops.

The online CF-related health outcomes survey will present consumers with the same two open-ended questions as those posed at the preliminary workshops (Appendix 1). This is being performed to ensure broad capture of CF-related health outcomes. The survey will be advertised and remain open for a 4-week period from commencement.

**Data collection instruments and technologies**

Workshops will be audio-recorded to enable playback, which is necessary to ensure the validity of data by minimising investigator recall bias.

CF-related health outcomes survey will be built using a REDCap online database, which will be hosted on a secure server at the Telethon Kids Institute. DCE will be built by a commercial provider. Both surveys will be conducted anonymously and will collect non-identifiable data only. Participants can exit from the online surveys at any time prior to submission of their responses. After this time, it will not be possible to withdraw their responses, as all items are non-identifiable.

**Data processing**

Workshop and causal diagram data files will be stored as password protected Excel or word documents. The non-identifiable CF-related health outcomes results data set will be downloaded from REDCap. The non-identifiable DCE data set will be sent as a password protected file by the commercial provider.

All data files will be stored securely on a password protected computer, which will be backed up on the Telethon Kids Institute server. Hard copy consent forms will be stored securely in a fireproof, locked filing cabinet at Telethon Kids Institute. The Institute is protected by high-level security and requires swipe card access for entry to the building and individual work areas. Data and research records will be retained for a minimum of 5 years after the date of last publication or until the youngest subject turns 25 years of age (whichever occurs later).

**Data analysis**

Analysis for the DCE will be performed in STATA V.13 using a range of regression approaches. For initial analysis, we will conduct a conditional logit. This will be used to understand the treatment preferences and trade-offs made by patients when considering outcomes relating to treatment of pulmonary exacerbations. For conditional logit analysis, the functional form is specified as:

\[ U_{ij} = \beta x_{ij} + \varepsilon_{ij} \]

which represents the utility of option \( j \) in choice set \( s \) for survey respondent \( i \), where \( x_{ij} \) is a vector of dummy variables representing the levels of the health state presented in option \( j \). \( \beta \) is a vector of utility weights associated with each level and \( \varepsilon_{ij} \) is the error term. \(^{36}\)

Second, we will use a mixed logit model to evaluate preference heterogeneity among respondents:

\[ U_{ij} = (\beta + n_{i})x_{ij} + \varepsilon_{ij} \]

where \( \beta \) represents population mean preferences and \( n_{i} \) is the individual deviation around those mean preferences. \(^{36}\)

Additionally, we will run exploratory analyses using a generalised multinomial logit, which considers both scale and preference heterogeneity. \(^{37, 38}\) However, this will not be the prespecified primary outcome as there is concern about its ability to converge with a relatively small sample size. An exploratory analysis on DCE responses will also be conducted using causal Bayesian networks. \(^{39}\) Causal Bayesian networks are a generalisation of the path models of structural equation modelling, \(^{22}\) which have been recently applied in DCE analysis to provide greater insight into choice processes. \(^{21}\)

**ETHICS AND DISSEMINATION**

Ethics approval for all aspects of this study was granted. Deviations from this protocol will not occur without prior approval. This study will be conducted in accordance with the International Council for Harmonisation of technical requirements for pharmaceuticals for human use (ICH) guidelines for Good Clinical Practice. \(^{40}\)

Participant information sheets will be provided to workshop and survey participants. Asking consumers to consider health-related outcomes may result in distress. Participants will be warned about this risk, and patients will be recommended to contact their general practitioner, CF clinic or Lifeline if this occurs.

Data obtained from workshop sessions or survey responses will remain confidential. Data will be reported in such a way that it will not be possible to identify individuals or their contributions.

Dissemination will occur through peer-reviewed publications and presentations to relevant stakeholders and research networks. DCE results will be reported according to the Guidance for Reporting Involvement of Patients and the Public checklist. \(^{41}\) This is a consensus reference document agreed by international representatives, which provides guidance about how to report patient and public involvement in health-related and social research.

**Author affiliations**

1Infectious Diseases, Perth Children’s Hospital, Nedlands, Western Australia, Australia
2School of Medicine, University of Western Australia, Nedlands, Western Australia, Australia
3School of Public Health, Curtin University, Bentley, Western Australia, Australia
4Respiratory Medicine, Perth Children’s Hospital, Nedlands, Western Australia, Australia
5Clayton School of IT, Monash University, Clayton, Victoria, Australia
6Intensive care, St John of God Hospital, Subiaco, Western Australia, Australia
7School of Population Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia

Acknowledgements The authors thank consumer advocates Mitch Messer and Kate Spaapen for their valued involvement in facilitating the consumer engagement activities relating to this body of work.

Contributors TS was responsible for the overall study concept. TS, CM, RN, AS, SM and SW elaborated the study protocol. CM drafted the manuscript. All authors revised and approved the final manuscript. All authors meet the ICMJE criteria for authorship.

Funding This work was supported by the Perth Children’s Hospital Foundation (New Investigator grant, 2017). Fellowship support was provided by the NHMRC (GNT1150996), the Wesfarmers Centre (top-up grant) and the Perth Children’s Hospital (Top-up grant).

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval All aspects of this study were approved by the Child and Adolescent Health Service Human Research Ethics Committee [RGS903].

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement There are no data in this work.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

REFERENCES

Chapter 7: A novel method to select outcomes for evaluation in trials and patient-centric outcome measures

7.1 Chapter summary

There is no consensus approach to selecting outcomes and endpoints when assessing the effects of interventions in clinical trials. Evaluation of outcomes and endpoints that are not causally related to the disease of interest or dependent on the intervention being studied may produce misleading information (Prentice, 1989; Fleming, 2012). Causal network diagrams may help clinical researchers identify outcomes that are both clinically meaningful and likely to be causally dependent on the intervention, and endpoints that are, in turn, causally dependent on those outcomes. We aimed to (i) develop an approach that can be applied across research domains for selecting outcomes and endpoints for evaluation in clinical trials and (ii) apply this to select a limited number of patient-centred outcomes for inclusion in two MAOI’s; one for evaluating the impact of treatments for pulmonary exacerbations of CF for use in adolescents and adults and the other for use in children with CF.

The following prioritised outcomes that independently related to the causal pathophysiological processes of disease were: painful/difficult breathing, sputum production/clearance, fatigue, poor appetite, pain, motivation/demoralisation, fevers/night sweats, treatment burden, inability to meet goals (personal, school, or work) and gastrointestinal symptoms (constipation, bloating and flatulence). Together, these outcomes collectively capture important aspects of the overall patient experience. These outcomes will be included in the MAOI’s; these are presented in Chapter 8.

7.2 Journal article


Supplementary materials:

Appendix 2: S1 Initial online outcomes elicitation survey
Appendix 3: S2 Prioritisation of outcomes survey
Appendix 4: S3 List of outcomes elicited at workshops and identified in the literature review
Appendix 5: S4 General subnetwork
Appendix 6: S5 Gastrointestinal subnetwork
Appendix 7: S6 Mental health subnetwork
Appendix 8: S7 Mental health subnetwork
A novel method to select meaningful outcomes for evaluation in clinical trials

Charlie McLeod\textsuperscript{a,b}\textsuperscript{*} E: charlie.mcleod@health.wa.gov.au, Richard Norman\textsuperscript{c} E: richard.norman@curtin.edu.au, Jamie Wood\textsuperscript{d,e} E: jamie.wood@mountsinai.org; Siobhain Mulrennan\textsuperscript{f,g} E: Siobhain.Mulrennan@health.wa.gov.au; Sue Morey\textsuperscript{j} E: Sue.Morey@health.wa.gov.au; André Schultz\textsuperscript{b,j} E: Andre.Schultz@health.wa.gov.au; Mitch Messer\textsuperscript{b} E: Mitch.Messer@telethonkids.org.au; Kate Spaapen, E: katerybarczyk@gmail.com; Matt Stoneham, E: matthew.stoneham1@bigpond.com; Yue Wu\textsuperscript{l} E: yue.wu@telthonkids.org.au; Alan R Smyth\textsuperscript{h} E: Alan.Smyth@nottingham.ac.uk, Christopher C. Blyth\textsuperscript{a,b} E: Christopher.blyth@uwa.edu.au, Steve Webb\textsuperscript{k,m} E: steve@stevewebb.com.au, Steven Mascaro\textsuperscript{n} E: steven.mascaro@bayesian-intelligence.com; Owen Woodberry\textsuperscript{o}, E: owen.woodberry@bayesian-intelligence.com; Thomas L. Snelling\textsuperscript{j,o} E: tom.snelling@sydney.edu.au.

\textsuperscript{a}Infectious Diseases Department, Perth Children’s Hospital, Nedlands, Australia
\textsuperscript{b}Wesfarmers Centre for Vaccines and Infectious Diseases, Telethon Kids Institute, Nedlands, Australia
\textsuperscript{c}School of Public Health, Curtin University, Bentley, Australia
\textsuperscript{d}Abilities Research Center, Department of Rehabilitation and Human Performance, Icahn School of Medicine at \textsuperscript{e}Physiotherapy Department, Sir Charles Gairdner Hospital, Nedlands, Australia
Mount Sinai, New York, United States of America
\textsuperscript{f}Respiratory Department, Sir Charles Gairdner Hospital, Nedlands, Australia
\textsuperscript{g}Faculty of Health and Medical Sciences, University of Western Australia, Crawley, Australia
Wal-yan Respiratory Research Centre, Telethon Kids Institute, University of Western Australia, Nedlands, Australia

Department of Respiratory Medicine, Perth Children’s Hospital, Nedlands, Australia

Sydney School of Public Health, The University of Sydney, Sydney, Australia

Evidence Based Child Health Group, School of Medicine, University of Nottingham, United Kingdom

Department of Intensive Care, St John of God Hospital, Subiaco, Australia

School of Population Health and Preventive Medicine, Monash University, St Kilda, Australia

Bayesian Intelligence, Upwey, Australia

Menzies School of Health Research, Royal Darwin Hospital Campus, Tiwi, Australia

*Corresponding author: Charlie McLeod; Infectious Diseases Department, Perth Children’s Hospital, 15 Hospital Avenue Nedlands, WA 6009; E: charlie.mcleod@health.wa.gov.au P: +61 8 6456 2222

Word Count: 3601
Abstract

Background:
A standardised framework for selecting outcomes for evaluation in trials has been proposed by the Core Outcome Measures in Effectiveness Trials working group. However, this method does not specify how to ensure that the outcomes that are selected are causally related to the disease and the health intervention being studied. Causal network diagrams may help researchers identify outcomes that are both clinically meaningful and likely to be causally dependent on the intervention, and endpoints that are, in turn, causally dependent on those outcomes. We aimed to (i) develop a generalisable method for selecting outcomes and endpoints in trials and (ii) apply this method to select outcomes for evaluation in a trial investigating treatment strategies for pulmonary exacerbations of cystic fibrosis (CF).

Methods:
We conducted a series of online surveys and workshops among people affected by CF. We used a modified Delphi approach to develop a consensus list of important outcomes. A workshop involving domain experts elicited how these outcomes were causally related to the underlying pathophysiological processes. Meaningful outcomes were prioritised based on the extent to which each outcome captured separate rather than common aspects of the underlying pathophysiological process.

Results:
The ten prioritised outcomes were: breathing difficulty/pain, sputum production/clearance, fatigue, appetite, pain (not related to breathing), motivation/demoralisation, fevers/night sweats, treatment burden, inability to meet personal goals and avoidance of gastrointestinal symptoms.

Conclusions:
This proposed method for selecting meaningful outcomes for evaluation in clinical trials may improve the value of research as a basis for clinical decisions.

**Keywords:**

Cystic Fibrosis, consumer, outcome assessment, patient reported outcomes, Bayesian network
Key messages:

What is the key question?

Which outcomes are both meaningful to people affected by CF and causally related to the underlying pathophysiological processes of disease and the mechanism of action of the intervention in question.

What is the bottom line?

This study identified ten outcomes that are both meaningful and likely to be causally affected by treatment(s) for pulmonary exacerbations in CF. This is the first step towards the development of weighted outcome measures for use when evaluating the effect of treatment interventions for pulmonary exacerbations in adults and children.

Why read on?

We present a rational approach for selecting outcomes that are meaningful, causally related to disease processes, and likely to be impacted by interventions under study; application of this method could improve the quality of clinical research.
1.0 Introduction

Selecting appropriate outcomes for evaluation in clinical studies is critical for ensuring the value of that research (1). The Core Outcome Measures in Effectiveness Trials (COMET) initiative was established in 2010, and provides a framework for the development of core outcome sets (COS). COS are a minimum set of meaningful outcomes that should be measured and reported in all trials of a specific disease or study population (2). COMET specifies that the outcomes included should arise from exposure to a causal factor or a health intervention (2). The rationale for this is that outcomes that are not causally related to the pathophysiology of the disease of interest and dependent on the mechanism of the intervention being studied may produce misleading information (3). Causal network diagrams may help researchers identify outcomes that are both clinically meaningful and likely to be causally dependent on the intervention, and endpoints that are, in turn, causally dependent on these outcomes; it may also help to identify endpoints that are likely to capture very similar information about the overall experience, because they are causally related to the same outcome, or to outcomes that are causally closely related (4).

Outcomes can be defined as patient characteristics or biological processes targeted for improvement by an intervention (e.g. lung function), and endpoints as the specific measurable parameter(s) corresponding to those outcomes (e.g. change in the percentage predicted forced expiratory volume in one-second [ppFEV$_1$] from baseline to day 10) [McLeod et al, unpub]. Outcomes in clinical studies should be patient-centred and meaningful; that is, they should capture either directly, or indirectly (as a valid surrogate), how a person feels, functions or survives (5). To this end, outcomes should be considered important to those affected by the disease in question. It’s increasingly recognised as
important to include consumers in the selection of outcomes in clinical studies, but this has only occasionally been done (5).

A causal network diagram (or directed acyclic graph) can help to represent the causal relationships between relevant factors (whether measurable or not), including the various relevant outcomes of an intervention for a given disease. Causal network diagrams use unidirectional arrows, or ‘arcs’, to connect factors in a pairwise fashion in which the direction of the arc represents the causal direction of the relationship, i.e. from cause to effect (6). They can be used to explicitly represent and explore our understanding of the causal mechanisms underlying a problem domain, or for clinical problems, the pathophysiological processes which give rise to various symptoms, functional outcomes and disease states (including death), and how and where in the process various treatments are thought to act (7).

Causal network diagrams may also help to identify factors that are not related causally to outcomes of clinical importance, and which are therefore unlikely to be useful or reliable as surrogate outcomes, even if correlated with the outcome (4, 8). For example, the number of cigarettes smoked per day may correlate strongly with lung function in people with emphysema; if evaluating the impact of a quit smoking intervention, measuring any reduction in smoking may be a useful surrogate for (eventual) improved lung function, but measuring reduction in smoking would not be reliable as a surrogate if evaluating the effect of inhaled steroid therapy on lung function in emphysema. Inspection of a causal network diagram (see Figure 1) quickly reveals that smoking intensity lies as a mediator on the causal pathway between the quit smoking intervention and lung function, as it is both an effect of the intervention and the cause of the change in lung function. But smoking intensity does not lie as a mediator on the causal pathway between steroid therapy and lung function, even if
smoking and lung function remain strongly correlated. Finally, causal network diagrams may help to identify any causal relationships between outcomes, particularly whether one outcome is causally dependent on another, or whether they share another outcome as a common cause. In general, sets of outcomes should be selected in a way that ensures they provide maximal information about the overall outcome for the patient. For example, if measuring the impact of a quit smoking intervention using two outcomes, its impact on dyspnoea and angina may be more informative than its impact on dyspnoea and cough (which share airway inflammation as a common cause), or its impact on angina and myocardial infarction (which share coronary atherosclerosis as a common cause).

![Causal diagram: smoking and myocardial infarction](image)

**Figure 1**  
*Causal diagram: smoking and myocardial infarction*

Here we present a novel approach for selecting meaningful patient-centred outcomes when evaluating pulmonary exacerbations in people with CF. We aimed to (i) develop a generalisable method for selecting outcomes and endpoints for evaluation in clinical studies and (ii) apply this method to select outcomes for use when evaluating interventions for pulmonary exacerbation in CF. Our ultimate goal is to include these outcomes in outcome
measure instruments which could be used to compare interventions by observing disparate endpoints among trial participants.

2.0 Methods

2.1 Overview

This project comprised four stages conducted between October 2018 and August 2019 (Figure 2). Written or online consent was obtained for participation in all stages. Ethics approval was provided by the Child and Adolescent Health Service Human Research Ethics Committee (RGS0000000903). Participants were not paid to participate, however those who attended workshops in-person received small compensation for incurred costs.

![Figure 2: Stages of Research](image)

2.2 Patient and public involvement

People affected by CF were involved in the design and conduct of this research. During the feasibility stage, formulation of the specific research questions, methods of recruitment and
wording for questionnaires were informed by workshop discussions with people affected by CF. Two consumers (MM and KS) were also included as investigators. This research area has also been identified as a priority for research by consumers through a previous James Lind consumer priority setting exercise (9). Once published, participants will be informed of the results through a dedicated website (https://adaptivehealthintelligence.org.au).

2.3 Stage 1

The literature was reviewed to identify a comprehensive range of outcomes and endpoints reported in CF clinical studies, and two 2-hour workshops were held using a modified Delphi approach (10) to elicit meaningful health outcomes among (i) people ≥13 years with CF and (ii) carers of children or adults with CF. We aimed to elicit answers to two questions: (i) ‘What CF-related health outcomes are important to people affected by CF?’ and (ii) ‘What adverse effects of treatment are important to people affected by CF?’ Attendees were asked to suggest their own outcomes, and when this was exhausted, they were asked to consider the importance of any other outcomes identified in the literature review. The workshop focussed explicitly on clinically meaningful outcomes, that is outcomes that aim to capture how a person feels or functions, rather than mechanistic outcomes (e.g. spirometry, radiographic changes, and sputum or blood inflammatory biomarkers). Workshops for carers were conducted in person, while those for people with CF were conducted via videoconference owing to the infection control risk of face-to-face contact. An online survey comprising the same two questions was conducted over 5 weeks (see SI); participation was voluntary and advertised via local and national CF consumer groups and research networks, and the main paediatric and adult CF treatment centres in Perth, Australia.
### 2.4 Stage 2

Two further workshops were conducted for similarly comprised groups and under the same conditions described for stage 1. The aim of these workshops was to obtain agreement on a prioritisation of the outcomes identified in stage 1, capturing how people feel and function from the perspective of (i) people with CF ≥13 years old and (ii) carers. The prioritisation exercise (see S2) was also conducted as an online survey over 3 weeks.

### 2.5 Stage 3

An expert knowledge elicitation workshop (facilitated by SM) was conducted to try to derive an agreed causal network diagram linking the outcomes identified in stages 1 and 2 to underlying pathophysiological processes, and to each other. A group of clinical domain experts in CF pulmonary exacerbations, people with CF ≥13 years old and carers attended. Experts were provided a brief explanation of causal networks and then asked to consider and either agree with, or revise, a baseline causal framework in which airway infection gives rise to inflammation (treatable by various agents) which in turn affects functional outcomes and gives rise to a range of symptoms. On agreement that this basic framework was appropriate and sufficient for the task, emphasis shifted to specific symptoms and functional outcomes. Prior to the workshop, these were partitioned into potentially discrete domains or subnetworks. Experts were provided the opportunity to review and revise these subnetworks, and were also allowed to adjust them as the workshop proceeded if the need became apparent.
Experts were asked to consider each such subnetwork in turn, and in isolation from other subnetworks. Initially, subnetworks consisted of two to ten factors that were fully disconnected. Experts were prompted to identify the most causally important factors and, in particular, key common causes that might strongly influence many other factors within the subnetwork. They were also advised to keep the subnetworks simple, omitting weak causal influences, and to avoid the creation of cycles (in which a factor is described as causally dependent on itself) by specifying only the most immediate and dominant direction of causal influences. As experts suggested causal relationships between factors, the facilitator added them to the network, and the group considered them in terms of their strength, redundancy or potential conflict with other relationships. Based on the group consensus, causal relationships were then either retained or removed. Any factors that remained without any significant causal relationships to other factors at the end of the process were re-considered and either retained, moved to another subnetwork or removed entirely if no longer considered worth keeping.

After each subnetwork was considered in isolation, the connections between subnetworks were considered in much the same way, with the exception that experts could suggest connections between either the subnetworks themselves or between any individual factors contained within them.

After the workshop, subnetworks were converted into Bayesian network submodels (structure only) and reviewed a second time by a smaller group of CF pulmonary exacerbation domain experts for any remaining issues and inconsistencies.

2.6 Stage 4
A single agreed list of ten priority outcomes across both people with CF ≥13 years old and carers was produced by representatives from the group above. Priority outcomes were selected based on two considerations; (i) the importance of the outcome as assessed by both groups AND (ii) the extent to which each outcome captured separate rather than common aspects of the underlying pathophysiological process and the overall outcome experience.

3.0 Results

3.1 Stage 1 & 2

Thirty-six people participated in stage 1: six people with CF ≥13 years old (one was <25 years old) and eight carers participated via workshop; seven people with CF (two aged <25 years old) and 15 carers participated via the online survey.

Fifty-five people participated in stage 2: 12 people with CF (one aged < 25 years old) and 15 carers attended the prioritisation workshops; and 12 people with CF (none aged < 25 years old) and 16 carers participated via the online survey.

A condensed list of identified outcomes based on results from stages 1 & 2 are presented in Table 1. A more extensive list is provided in S3.
<table>
<thead>
<tr>
<th>Rank</th>
<th>People with CF ≥13 years</th>
<th>Carers of children 0-18 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Coughing up blood</td>
<td>Coughing up blood</td>
</tr>
<tr>
<td>2</td>
<td>Shortness of breath/difficulty breathing</td>
<td>Shortness of breath/difficulty breathing</td>
</tr>
<tr>
<td>3</td>
<td>Feeling fatigued/deconditioned</td>
<td>Presence of pain</td>
</tr>
<tr>
<td>4</td>
<td>Anxiety/worry</td>
<td>Feeling anxious/worried</td>
</tr>
<tr>
<td>5</td>
<td>Sputum production (presence of, or worsening from baseline)</td>
<td>Feeling sad/depressed</td>
</tr>
<tr>
<td>6</td>
<td>Sadness/depressed mood</td>
<td>High treatment burden</td>
</tr>
<tr>
<td>7</td>
<td>Tiredness</td>
<td>Gastrointestinal (abdominal pain, diarrhoea, flatulence)</td>
</tr>
<tr>
<td>8</td>
<td>High treatment burden</td>
<td>Impaired hearing</td>
</tr>
<tr>
<td>9</td>
<td>Inability to cough/clear up sputum</td>
<td>Sputum production (presence of, or worsening from baseline)</td>
</tr>
<tr>
<td>10</td>
<td>Inability to meet personal/school/work goals</td>
<td>Poor appetite/eating difficulties</td>
</tr>
<tr>
<td>11</td>
<td>Feeling unwell</td>
<td>Chest tightness</td>
</tr>
<tr>
<td>12</td>
<td>Poor exercise tolerance</td>
<td>Lack of energy</td>
</tr>
<tr>
<td>13</td>
<td>Nausea/vomiting</td>
<td>Not feeling well</td>
</tr>
<tr>
<td>14</td>
<td>Coughing</td>
<td>Reduced exercise tolerance</td>
</tr>
<tr>
<td>15</td>
<td>Reduction of usual activities</td>
<td>Fever</td>
</tr>
<tr>
<td>16</td>
<td>Nausea/vomiting</td>
<td>Inability of child to meet personal/school/work goals</td>
</tr>
<tr>
<td>17</td>
<td>Difficulty sleeping</td>
<td>Poor weight</td>
</tr>
<tr>
<td>18</td>
<td>Presence of pain</td>
<td>Difficulty sleeping</td>
</tr>
<tr>
<td>19</td>
<td>Presence and severity of wheezing</td>
<td>Reduced ability to concentrate/think clearly</td>
</tr>
<tr>
<td>20</td>
<td>Being worried about your CF</td>
<td>Wheeze</td>
</tr>
<tr>
<td>21</td>
<td>Reduced ability to concentrate/think clearly</td>
<td>Presence of sweats/chills</td>
</tr>
<tr>
<td>22</td>
<td>Chest tightness</td>
<td>Reduction of usual activities</td>
</tr>
<tr>
<td>23</td>
<td>Coughing</td>
<td>Amount of school/work missed by child</td>
</tr>
<tr>
<td>24</td>
<td>Gastrointestinal (diarrhoea/bloating/flatulence)</td>
<td>Presence of headaches</td>
</tr>
<tr>
<td>25</td>
<td>Fever</td>
<td>Feeling unwell</td>
</tr>
<tr>
<td>26</td>
<td>Amount of school/work missed</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Poor weight gain or weight loss</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Presence of severity of chills/sweats</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>Irritable/feeling cranky</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 Patient-reported outcomes ranked by perceived importance by people ≥13 years affected by CF and carers of children
3.2 **Stage 3**

Ten domain experts in CF pulmonary exacerbation participated in stage 3, including two paediatric respiratory physicians, two adult respiratory physicians, two paediatric infectious diseases physicians, three adults with CF and one carer. Based on the outcomes identified in stages 1 and 2, participants settled on eight broad pathophysiological domains which became the focus of separate causal subnetworks: respiratory, gut, sinus, endocrine, hearing, mental health, general, and functional. General outcomes included those that captured an aspect of an individuals’ overall health, such as fatigue or appetite. Functional outcomes comprised those that related to an individuals’ ability to perform activities of daily living or realise their own aspirations.

*Figure 3* illustrates the relationship between outcome domains and *Figure 4* illustrates consensus causal subnetwork models for the respiratory domain. Causal subnetwork models for general symptoms, the gastrointestinal system, mental health and functional outcomes are included in S4-7. Lighter arrows were used to indicate domains that were causally related to all other nodes.
Figure 3  Causal subnetwork model demonstrating the relationships between outcome domains

Figure 4  Respiratory subnetwork model illustrating causal relationships

3.3 Stage 4
Endocrine, hearing and sinus-related outcomes were not as highly prioritised by people with CF or carers compared with other domains (see Table 1); individual outcomes within these domains were consequently not further considered for inclusion in the final list of priority outcomes.

Excessive sputum production/poor clearance was identified as an important respiratory outcome, and was found to be causally related to all other outcomes within the respiratory domain including severe cough, coughing ‘spasms’, wheezing, chest tightness, and haemoptysis. Breathing difficulty (described as shortness of breath or consciousness or awareness of breathing) was found to be important to people with CF and carers alike, but being a relatively uncommon symptom in children may be less applicable in studies of pulmonary exacerbations compared to alternatives such as sputum production/poor clearance. Unlike breathing difficulty, sputum production/poor clearance is also causally related to faecal and urinary incontinence via its relationship with coughing spasms, and so may capture additional useful information compared to breathing difficulty alone.

In the general symptoms domain, fevers/night sweats and poor appetite were identified as priority outcomes because they were meaningful and important to both people with CF and carers, and were relatively independent of the other outcomes already identified.

Pain was a highly prioritised as an outcome by carers and was considered moderately important to people living with CF; this outcome featured as important across a number of domains, including the general symptoms, respiratory and gastrointestinal domains. Pain emerged as a causally dominant outcome within the gastrointestinal domain. Other gastrointestinal symptoms (such as flatulence, diarrhoea and steatorrhoea) were also
identified as priorities, as they were meaningful outcomes among both people with CF and to carers, and were relatively causally independent of gastrointestinal pain.

In the mental health domain, a person’s overall feeling of amotivation/demoralisation was identified as a priority outcome, as it was itself causally influenced by a range of other important outcomes, including anxiety/worry, sadness or depression, irritability, and alcohol and drug dependence.

In the function domain, treatment burden and an inability to meet personal, school or work goals were identified as priorities; these outcomes were also found to be causally related to most other important outcomes in the functional sub-network, either as a common cause, or as a common effect of those outcomes. While not a functional outcome per se, it is worth noting that hospitalisation was found to be important given its impact on quality of life (QoL); it was also a factor that impacted on all outcomes in the function sub-network.

The ten outcomes prioritised by people with CF and carers that independently map to causal disease processes are presented in Figure 9.
4.0 Discussion

This is the first attempt to prioritise outcomes for reporting in studies of CF pulmonary exacerbations, as nominated by people with CF and their carers. Selection of these patient-centred outcomes was achieved using a novel approach. We used causal network diagrams to select a subset of ten meaningful outcomes that collectively capture as much of the overall outcome experience as possible. The next step will be to work with people with CF and carers to quantify the relative importance that they place on each of these outcomes. This weighting will inform the development of separate weighted outcome measurement instruments for use in children and adults with CF, as a single summary measure of the overall outcome of pulmonary exacerbations of CF, and which could therefore be used to evaluate interventions.

Research priorities for CF from the perspective of more than 1000 consumers from 23 countries have been reported (11), as well as an assessment of the extent to which current CF
studies match these priorities (12). However, there is not yet consensus on a core outcome set (COS) for universal adoption when evaluating interventions in studies of pulmonary exacerbations in CF.

People with CF and carers alike prioritised breathing difficulty, excessive sputum production/poor clearance, fatigue, pain, amotivation/demoralisation, high treatment burden and inability to meet personal/school/work goals as important and meaningful. Parents/carers gave higher priority than people with CF to hearing impairment, reduced appetite and gastrointestinal symptoms; conversely people with CF gave higher priority than carers to tiredness, sleeping difficulty, and ‘feeling unwell’. Failure to separately derive priority outcomes for these two groups could be a limitation, but we decided on this pragmatic approach (i) because of the considerable overlap in outcomes between the groups and the critical importance of carers, in particular, for young children, and (ii) because the intention was to use a causal understanding of the same underlying disease process, and people with CF, carers, and other domain experts all have valid expertise to contribute.

Limitations of this work included the relatively small number of participants in the consumer workshops; however almost all outcomes identified by review of the literature were also independently identified and considered by the participants, so we are confident that we have not overlooked important outcomes. Participants were largely from Australia where people have access to universal healthcare; this may limit the generalisability of our results to populations who may have poorer access to health care. While we assume that the views of participants in this study represent a broader population affected by CF, the age and disease severity of individual participants with CF is likely to materially impact on the ascertainment and perceived importance of each outcome. Young people with CF (aged < 25 years old)
were not well represented in either the workshops or the online surveys, and we did not try to separately ascertain priorities from children, or how their perspectives differ from that of their carer(s), which means we were unable to draw meaningful conclusions about the differences in outcome selection between different age groups. It is not entirely clear why it was difficult to recruit adolescents to this study. Possible contributing factors include the competing demands on their time or a lack of understanding about the relevance or importance of the study for them.

The framework we present here for selecting outcomes for use in clinical studies benefits from direct involvement of patients and families, and is arguably more rigorous than alternative methods which ignore how such outcomes causally relate to the underlying disease process and with each other. We expect that standardised reporting of these outcomes in clinical studies evaluating treatments for pulmonary exacerbations in CF would improve the value of those studies. We believe that this approach has the potential for broader application to the selection of outcomes in studies across various clinical problem domains, especially where the underlying pathophysiology is complex and the potential range of outcomes is broad.

Further work is required to engage children and adolescents living with CF to ascertain what health outcomes they prioritise, and how these preferences compare to those of their carer(s). It will also be necessary to quantify the relative importance of each of the outcomes identified here. A discrete choice experiment will be the focus of subsequent research; participants affected by CF will be presented with hypothetical choice sets and asked to choose between treatment options to ascertain how they value different outcomes resulting from treatment. Weighted outcomes will then be incorporated into a multi-attribute utility instrument.
designed to capture the impact of trial interventions as a single score. This tool will require validation as an outcome assessment tool and will be compared to traditional outcomes such as forced expiratory volume in one-second (FEV₁) before being considered for use more broadly.

**Funding**

This work was supported by a Perth Children’s Hospital New Investigator grant (9757). CM is supported by an NHMRC post-graduate scholarship (GNT1150996), and top-up grants from the Wesfarmers Centre (top-up grant) and the Perth Children’s Hospital Foundation (9722). TS is supported by a Career Development Fellowship from the National Health and Medical Research Council (GNT1111657). CB is also supported by the National Health and Medical Research Council (1111596/1173163).

**Acknowledgements**

The authors would like to thank the people with CF and carers of children less than 18 years of age with CF who were involved in the design, conduct, participation, reporting and dissemination of the results of this research. We would additionally like to thank Sophie Longton, Adelaide Withers, Anna Tai and David Armstrong for their contribution to the Bayesian Expert Knowledge Elicitation Workshop, and members of the Adaptive Health Intelligence group for their support in helping to facilitate the consumer workshops; particularly Inge Timmerman, Nelly Newall and Grace Currie.

**Conflicts of interest**

Andre Schultz receives an honorarium from Vertex as a member of the advisory board for work that is not related to this article.
References


Chapter 8: Preferred health outcomes following treatment for pulmonary exacerbations of CF

8.1 Chapter summary

This chapter presents the results of a discrete choice experiment (DCE) administered to a sample of adolescents and adults with CF and carers of children with CF. In the DCE survey, participants were asked to make a series of hypothetical treatment decisions to evaluate how they make trade-offs between different aspects (attributes) of health when undergoing treatment for pulmonary exacerbations. Data were analysed using a conditional logit regression model. Weights for attributes were derived from these data to enable calculation of a single weighted score between 0 (worst outcome state) and 100 (best outcome state).

Data from 362 participants (167 people 13 years and above with CF and 195 carers of children with CF) were included in the DCE analysis. Most respondents were female and resided in Australia. The key finding was that difficult/painful breathing had the greatest impact overall on the preferred health state among people with CF and carers alike. Avoidance of gastrointestinal problems also heavily influenced decision-making. The impact of treatment on study/work was considered important to adults with CF, however pain (unrelated to breathing) and anxiety/worry had a greater impact on decision-making from the perspective of carers. The impact of treatment getting in the way of ‘things you like to do’ had a stronger influence on the preferred health state of people with CF compared to carers whose preferences appeared to be influenced more by the presence of pain (unrelated to breathing) and anxiety/worry experienced by children under their care.

This is the first study to report patient preference information regarding treatment of pulmonary exacerbations of CF. These data should help inform clinical decision-making. The patient and proxy carer-reported weighted outcome measure instruments developed in this study will be validated using the BEAT CF trial cohort. Future studies should focus on obtaining preference information from children, including adolescents, and exploring how these preferences compare to those of their carer(s).

8.2 Submitted journal article

treatment for pulmonary exacerbations of cystic fibrosis (following page). *Journal of cystic fibrosis.* Submitted 6.5.21.

**Supplementary materials:**

Appendix 9: S1 Example DCE choice set from survey for people with CF  
Appendix 10: S2 Attributes and levels for the DCE survey for people with CF  
Appendix 11: S3 Attributes and levels for the DCE survey for carers  
Appendix 12: S4 DCE survey- people with CF  
Appendix 13: S5 DCE survey- adolescents with CF  
Appendix 14: S6 DCE survey- carers  
Appendix 15: S7 Retention of participants for DCE survey  
Appendix 16: S8 Survey feedback from participants
Preferred health outcome states following treatment for pulmonary exacerbations of cystic fibrosis

Charlie McLeod\textsuperscript{a-c,*} E: charlie.mcleod@health.wa.gov.au, Jamie Wood,\textsuperscript{d-e} E: jami.wood@mountsinai.org; Siobhain Mulrennan\textsuperscript{f-g} E: Siobhain.Mulrennan@health.wa.gov.au; Sue Morey\textsuperscript{f} E: Sue.Morey@health.wa.gov.au; André Schultz\textsuperscript{b,h-i} E: Andre.Schultz@health.wa.gov.au; Mitch Messer\textsuperscript{a} E: Mitch.Messer@telethonkids.org.au; Kate Spaapen, E: katerybarczyk@gmail.com; Yue Wu\textsuperscript{j} E: yue.wu1@sydney.edu.au; Steven Mascaro\textsuperscript{k} E: steven.mascaro@bayesian-intelligence.com; Alan R Smyth\textsuperscript{l} E: Alan.Smyth@nottingham.ac.uk, Christopher C. Blyth\textsuperscript{b-c,m-n} E: Christopher.blyth@uwa.edu.au, Steve Webb\textsuperscript{o,p} E: steve@stevewebb.com.au, Thomas L Snelling\textsuperscript{l-d} E: tom.snelling@sydney.edu.au, Richard Norman\textsuperscript{f} E: richard.norman@curtin.edu.au

\textsuperscript{a}Infectious Diseases Implementation Research Division, Telethon Kids Institute, Nedlands, 6009, Australia

\textsuperscript{b}Division of Paediatrics, Faculty of Medicine, University of Western Australia, Crawley, 6009, Australia

\textsuperscript{c}Infectious Diseases Department, Perth Children’s Hospital, Nedlands, 6009, Australia

\textsuperscript{d}Abilities Research Center, Department of Rehabilitation and Human Performance, Icahn School of Medicine at Mount Sinai, New York, NY 10029, United States of America

\textsuperscript{e}Respiratory Department, Sir Charles Gairdner Hospital, Nedlands, 6009, Australia

\textsuperscript{f}Faculty of Health and Medical Sciences, University of Western Australia, Crawley, 6009, Australia

\textsuperscript{g}Wal-yan Respiratory Research Centre, Telethon Kids Institute, University of Western Australia, Nedlands, 6009, Australia

\textsuperscript{h}Department of Respiratory Medicine, Perth Children’s Hospital, Nedlands, 6009, Australia
Sydney School of Public Health, The University of Sydney, Sydney, 2052, Australia

Bayesian Intelligence, Upwey, 3158, Australia

Evidence Based Child Health Group, School of Medicine, University of Nottingham, NG7 2RD, United Kingdom

Wesfarmers Centre for Vaccines and Infectious Diseases, Telethon Kids Institute, Nedlands, 6009, Australia

Pathwest Laboratory Medicine WA, QEII Medical Centre, Nedlands, 6009, Australia

Department of Intensive Care, St John of God Hospital, Subiaco, 6008, Australia

School of Population Health and Preventive Medicine, Monash University, St Kilda, 3004, Australia

Menzies School of Health Research, Royal Darwin Hospital Campus, Tiwi, 0810, Australia

School of Population Health, Curtin University, Bentley, 6102, Australia

*Corresponding author: Charlie McLeod; Infectious Diseases Department, Perth Children’s Hospital, 15 Hospital Avenue, Nedlands, 6009, Australia; E: charlie.mcleod@health.wa.gov.au P: +61 8 6456 2222

Running title: Preferred health outcome states following treatment for pulmonary exacerbations

Keywords: cystic fibrosis, discrete choice experiment, patient preferences, patient-reported outcome measures, outcomes
Abstract

*Background:* Treatment for pulmonary exacerbations of cystic fibrosis (CF) can produce a range of positive and negative outcomes. Understanding which of these outcomes are achievable and desirable to people affected by disease is critical to agreeing to goals of therapy and determining endpoints for trials. The relative importance of outcomes resulting from treatment of these episodes are not reported. We aimed to (i) quantify the relative importance of outcomes resulting from treatment for pulmonary exacerbations and (ii) develop patient and proxy carer-reported weighted outcome measures for use in adults and children, respectively.

*Methods:* An online discrete choice experiment (DCE) survey was conducted. Participants were asked to make a series of hypothetical decisions about treatment for pulmonary exacerbations to assess how they make trade-offs between different attributes of health. Data were analysed using a conditional logistic regression model. The correlation coefficients from these data were rescaled to enable generation of a composite health outcome score between 0 and 100 (worst to best health state).

*Results:* 362 individuals participated (167 people with CF and 195 carers); of these, 206 completed the survey (56.9%). Most participants were female and resided in Australia. Difficult/painful breathing had the greatest impact on the preferred health state among people with CF and carers alike. Avoidance of gastrointestinal problems also heavily influenced decision-making.

*Conclusions:* These data should be considered when making treatment decisions and determining endpoints for trials. Further research is recommended to quantify the preferences of children and whether these align with those of their carer(s).
1.0 Introduction

There are no data characterising the relative importance of outcomes resulting from treatment of pulmonary exacerbations of cystic fibrosis (CF) from the perspective of people affected by disease. Pulmonary exacerbations are a hallmark of disease and are typically characterised by acute worsening of respiratory symptoms and deterioration in lung function (1). Management of these episodes is complex and generally involves a combination of antimicrobial, anti-inflammatory and muco-active agents, chest physiotherapy and optimisation of nutrition (2). These treatments target different aspects of the disease process, and may therefore variably affect different outcomes. Moreover, treatments differ with respect to toxicities and burden on people with CF and their carers; adverse consequences may need to be traded off against potential benefits when considering treatment options. Treatment decisions should arguably reflect the shared goals of therapy based on outcomes that are considered achievable and desirable to both people with CF (and their carers where appropriate) and health professionals.

A variety of methods exist for eliciting patient preference information. A discrete choice experiment (DCE) is one method for quantifying how people value different aspects of health-related status (3). A DCE is designed based on three key principles; (i) participants are asked to choose between alternatives when confronted with realistic, hypothetical choice set scenarios described by various attributes (outcomes), (ii) the choice between alternatives depends on the attribute levels and (iii) choices are based on a latent utility function (4). DCEs differ from other stated preference methods because participants are forced to weigh the relative importance of attributes (such as drug benefit versus toxicity), compared to other techniques which simply rank or rate them (3). DCEs are being increasingly used as a method for generating weights to inform the development of multi-attribute instruments to quantify the value of different health states (5).
To minimise the cognitive burden on participants, it is necessary to restrict the number of attributes studied in DCEs (3, 5, 6). We previously systematically reviewed outcomes reported in trials for pulmonary exacerbations of CF (7) and conducted a series of surveys and workshops to elicit outcomes of importance from the perspective of people affected by CF. A separate workshop involving health professionals and people affected by CF explored how these outcomes were causally dependent on underlying pathophysiological disease processes. A consensus set of ten prioritised outcomes (capturing symptoms or function) for inclusion in the DCE was derived based on two considerations; (i) their importance from the perspective of people affected by disease and (ii) the extent to which each outcome captured distinct rather than similar or closely related outcomes. The ten outcomes were breathing difficulty/pain, sputum production and clearance, fatigue, appetite, pain unrelated to breathing, motivation/demoralisation, fevers/night sweats, treatment burden, inability to meet goals and avoidance of gastrointestinal symptoms [McLeod, accepted for publication].

Here we present the results of a DCE evaluating the preferred outcomes resulting from treatment of pulmonary exacerbations from the perspective of adolescents and adults with CF (hereafter referred to as people with CF) and carers of children with CF (hereafter referred to as carers). The DCE was performed to inform weights for the attributes included in two weighted outcome measure instruments (hereafter referred to as instruments); one for use in people with CF and one for use in children with CF. It is anticipated that these instruments will be used to facilitate comparisons of treatment strategies in future clinical trials by providing a single weighted score capturing the overall patient experience.

2.0 Methods

The protocol for development of the DCE is reported elsewhere (8). Ethics approval was granted by the Child and Adolescent Health Service Human Research Ethics Committee (RGS903). Separate
versions of the DCE were produced for people with CF and for carers, however similar attributes were evaluated in both. The DCE comprised 12 choice set tasks. For each task, participants were instructed to choose between two hypothetical treatment options, where the levels of attributes can be different. We presented choice tasks where five attributes differed between the two alternatives to make the task easier for participants. See SI for an example choice set task. To further reduce the cognitive burden for participants, we highlighted those dimensions that differed between the treatments following the approach of Norman et al (9).

2.1 Selection of attributes and levels
The ten prioritised outcomes previously identified were included as the attributes studied in the DCE [McLeod, accepted for publication]. We consulted people affected by CF to inform the selection of wording used in the DCE and to set levels for attributes which were both meaningfully different and thought to represent the spectrum of clinical possibilities (S2-S3). One notable difference was the wording used to capture motivation/demoralisation between versions. For people with CF, the outcome thought to best capture motivation/demoralisation was ‘how often you feel overwhelmed’ compared to ‘how often your child feels sad or worried’ for carers.

2.2 Experiment and questionnaire design
An online survey was built using a platform created in Survey Engine (10). The survey was designed using Ngene V1.2 (ChoiceMetrics, Australia). The survey comprised five sections (see S4-6); (1) participant information and consent (2) an animated example choice set task, (3) demographic information, (4) 12 choice sets and (5) participant feedback.

The following criteria were used to guide the design of the DCE; (i) all levels and combinations must be reasonable, (ii) levels and combinations must be familiar to participants and (iii) levels should be sufficiently heterogeneous such that a trade-off is required when deciding between alternatives (3). A fractional factorial design was used; blocks of 12 choice sets derived from a pool
of 200 choice sets were randomly allocated to prospective participants. Choice sets were checked for plausibility, including by three consumers (one adolescent and one adult with CF and one carer).

Participant remuneration was not initially permitted by the approving ethics committee but was subsequently allowed providing it was not advertised. At the completion of the survey, AUD $20/gift card equivalent was offered as direct compensation or was committed as a donation to CF Australia instead.

2.3 Recruitment and eligibility

Recruitment occurred at the Perth Children’s Hospital and Sir Charles Gairdner Hospital in Australia and via consumer and research networks within Australia and the United Kingdom, using various methods including in-person recruitment and via electronic/social media. People 13 years and older with CF and carers of children less than 18 years old with CF were eligible. Participants less than 18 years old were also required to provide guardian consent (see S5).

2.4 Data analysis

Analyses for the DCE were specified a priori (11) and performed in STATA 13. A separate analysis for adolescents and adults with CF was originally intended, however owing to the small number of adolescent participants, these data were combined. For the initial analysis, a conditional logistic regression model was used. This was performed to understand the treatment preferences and trade-offs made by participants when considering outcomes relating to treatment. For analysis, the following functional form was specified:

\[ U_{isj} = \beta x_{isj} + \epsilon_{isj}, \]

which represents the utility of option \( j \) in choice set \( s \) for survey participant \( i \), where \( x_{isj} \) was a vector of dummy variables representing the levels of the health state presented in option \( j \), \( \beta \) was a vector of utility weights associated with each level, and \( \epsilon_{isj} \) was the error term. We used clustered standard errors which relaxes the assumption of independence between choice
sets. The levels of each attribute were designed to be ordered from least to most severe. We imposed this structure on the model. To do this, we first estimated an unconstrained model. Where mis-orderings occurred, we combined levels to prevent this, and re-estimated the model. The size of the combined coefficient was used to infer the strength of influence of each attribute level on the overall health state preference of participants.

To convert the regression results into an index, we rescaled the composite scores for individual health states such that the best possible health state (level 0 on each attribute) was scored at 100, and the worst possible health state (level 3 on each attribute) was scored at 0. As this scale is not anchored using dead as a health state, these figures cannot be used for constructing quality-adjusted life years, but do represent a weighted preference index which captures the relative value of different aspects of exacerbation-related outcomes in CF.

3.0 Results

Recruitment occurred between the 7th of May 2020 and the 15th of December 2020. Overall, 362 individuals consented to participate, of whom 156 were adults (43.1%) and 11 were adolescents (3.0%) with CF and 195 were carers (53.9%). Of those who consented, 119 people with CF (71.3%) and 139 carers (71.3%) responded to the first choice set, and 99 people with CF (59.3%) and 107 carers (54.9%) completed all choice set tasks. Of those who completed the survey, the median time to completion was 9.5 minutes for people with CF [IQR 7.4-13.4] and 12.4 minutes for carers [IQR 8.3-17.5]. See S7 for further details regarding the retention of participants in the survey.

3.1 Participant characteristics

Table 1 summarises the characteristics of participants. Most participants were women. The age of participants ranged from 13 to 63 years. Participants predominantly resided in Australia. Carers
reported fewer pulmonary exacerbations for the children under their care compared to people affected by disease, and better general health status.

<table>
<thead>
<tr>
<th>Gender (M:F)*</th>
<th>People with CF (n=162 unless otherwise stated, %)</th>
<th>Carers (n=186 unless otherwise stated, %)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>58 (35.8)</td>
<td>22 (11.8)</td>
</tr>
<tr>
<td>Median age, years [IQR]*</td>
<td>35 [26-45]</td>
<td>40 [35-48] (n=185)</td>
</tr>
<tr>
<td>Country of residence*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>150 (92.6)</td>
<td>180 (96.8)</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>9 (5.6)</td>
<td>1 (0.01)</td>
</tr>
<tr>
<td>United States of America</td>
<td>2 (1.2)</td>
<td>2 (0.01)</td>
</tr>
<tr>
<td>Italy</td>
<td>1 (0.6)</td>
<td>-</td>
</tr>
<tr>
<td>New Zealand</td>
<td>-</td>
<td>1 (0.01)</td>
</tr>
<tr>
<td>South Africa</td>
<td>-</td>
<td>1 (0.01)</td>
</tr>
<tr>
<td>Switzerland</td>
<td>-</td>
<td>1 (0.01)</td>
</tr>
<tr>
<td>Highest level of education of respondent*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary school completed</td>
<td>7 (4.3)</td>
<td>5 (2.7)</td>
</tr>
<tr>
<td>Secondary school completed</td>
<td>45 (27.8)</td>
<td>25 (13.4)</td>
</tr>
<tr>
<td>Tertiary education</td>
<td>48 (29.6)</td>
<td>77 (41.4)</td>
</tr>
<tr>
<td>Higher degree by research</td>
<td>10 (6.2)</td>
<td>25 (13.4)</td>
</tr>
<tr>
<td>Trade/diploma/other</td>
<td>52 (32.1)</td>
<td>54 (29.0)</td>
</tr>
<tr>
<td>Perceived health status**</td>
<td>(n=161)</td>
<td>(n=187)</td>
</tr>
<tr>
<td>Excellent</td>
<td>11 (6.8)</td>
<td>50 (26.7)</td>
</tr>
<tr>
<td>Very good</td>
<td>50 (31.1)</td>
<td>67 (35.8)</td>
</tr>
<tr>
<td>Good</td>
<td>59 (36.6)</td>
<td>54 (28.9)</td>
</tr>
<tr>
<td>Fair</td>
<td>37 (23.0)</td>
<td>11 (5.9)</td>
</tr>
<tr>
<td>Poor</td>
<td>4 (2.5)</td>
<td>5 (2.7)</td>
</tr>
<tr>
<td>Pulmonary exacerbations in the past 12 months**</td>
<td>(n=161)</td>
<td>(n=187)</td>
</tr>
<tr>
<td>None</td>
<td>64 (39.8)</td>
<td>120 (64.2)</td>
</tr>
<tr>
<td>1-2</td>
<td>77 (47.8)</td>
<td>50 (26.7)</td>
</tr>
<tr>
<td>3-4</td>
<td>17 (10.6)</td>
<td>8 (4.3)</td>
</tr>
<tr>
<td>&gt;5</td>
<td>3 (1.9)</td>
<td>9 (4.8)</td>
</tr>
</tbody>
</table>

*For the respondent **For carers, this question refers to the child under their care

| 3.2 Results of the main effect model: people with CF |

Among people with CF, the coefficients for the attribute levels generally reflected the expected monotonic relationship, meaning that the strength of the coefficients for attributes influencing the preferred health state were higher for levels that appeared worse at face value. For six attributes, levels were combined (Table 2). The five most influential attributes on the preferred health states were (1) difficulty/painful breathing (2) avoidance of gastrointestinal symptoms (3) fevers/night sweats, (4) how often treatment got in the way of ‘things you like to do’ and (5) mucus production/clearance. Coughing up mucus was important to participants only if it occurred ‘a lot’ or in ‘huge amounts.’ Treatment getting in the way of ‘things you like to do’ or how much treatment
impacted the ‘ability to keep up with school/studies/work’ was important if present ‘often’ or ‘always’. Poor appetite at meal times and pain unrelated to breathing had the least impact overall on preferences.

<table>
<thead>
<tr>
<th>Attribute/level</th>
<th>Raw coefficients</th>
<th>P-value</th>
<th>Combined coefficients</th>
<th>P-value</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>How often it's hard or painful to breathe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>0.000</td>
<td>-</td>
<td>0.000</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sometimes</td>
<td>-0.599</td>
<td>0.004</td>
<td>-0.594</td>
<td>0.003</td>
<td>-0.985 -0.204</td>
</tr>
<tr>
<td>Often</td>
<td>-0.867</td>
<td>0.000</td>
<td>-0.773</td>
<td>0.000</td>
<td>-1.147 -0.399</td>
</tr>
<tr>
<td>Always</td>
<td>-1.440</td>
<td>0.000</td>
<td>-1.491</td>
<td>0.000</td>
<td>-1.979 -1.002</td>
</tr>
<tr>
<td>How much mucus you cough up/swallow</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>0.000</td>
<td>-</td>
<td>0.000</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A little</td>
<td>0.161</td>
<td>0.362</td>
<td>-0.000</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A lot</td>
<td>-0.112</td>
<td>0.488</td>
<td>-0.185</td>
<td>0.202</td>
<td>-0.470 0.100</td>
</tr>
<tr>
<td>Huge amounts</td>
<td>-0.568</td>
<td>0.003</td>
<td>-0.636</td>
<td>0.000</td>
<td>-0.959 -0.312</td>
</tr>
<tr>
<td>How often you feel tired or lacking in energy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>0.000</td>
<td>-</td>
<td>0.000</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sometimes</td>
<td>0.176</td>
<td>0.247</td>
<td>-0.000</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Often</td>
<td>-0.026</td>
<td>0.879</td>
<td>-0.100</td>
<td>0.516</td>
<td>-0.402 0.202</td>
</tr>
<tr>
<td>Always</td>
<td>-0.247</td>
<td>0.275</td>
<td>-0.278</td>
<td>0.156</td>
<td>-0.662 0.106</td>
</tr>
<tr>
<td>How often you don't feel like eating at meal times</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>0.000</td>
<td>-</td>
<td>0.000</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sometimes</td>
<td>-0.011</td>
<td>0.950</td>
<td>-0.004</td>
<td>0.981</td>
<td>-0.305 0.297</td>
</tr>
<tr>
<td>Often</td>
<td>0.086</td>
<td>0.673</td>
<td>-0.004</td>
<td>0.981</td>
<td>-0.305 0.297</td>
</tr>
<tr>
<td>Always</td>
<td>-0.312</td>
<td>0.070</td>
<td>-0.344</td>
<td>0.034</td>
<td>-0.662 -0.025</td>
</tr>
<tr>
<td>How often you feel pain or discomfort (not related to breathing)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>0.000</td>
<td>-</td>
<td>0.000</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sometimes</td>
<td>-0.048</td>
<td>0.814</td>
<td>-0.004</td>
<td>0.977</td>
<td>-0.290 0.282</td>
</tr>
<tr>
<td>Often</td>
<td>0.113</td>
<td>0.528</td>
<td>-0.004</td>
<td>0.977</td>
<td>-0.290 0.282</td>
</tr>
<tr>
<td>Always</td>
<td>-0.454</td>
<td>0.004</td>
<td>-0.454</td>
<td>0.003</td>
<td>-0.752 -0.157</td>
</tr>
<tr>
<td>How often you feel overwhelmed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>0.000</td>
<td>-</td>
<td>0.000</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sometimes</td>
<td>-0.052</td>
<td>0.808</td>
<td>-0.035</td>
<td>0.861</td>
<td>-0.424 0.355</td>
</tr>
<tr>
<td>Often</td>
<td>-0.134</td>
<td>0.460</td>
<td>-0.178</td>
<td>0.288</td>
<td>-0.507 0.151</td>
</tr>
<tr>
<td>Always</td>
<td>-0.369</td>
<td>0.042</td>
<td>-0.392</td>
<td>0.022</td>
<td>-0.727 -0.057</td>
</tr>
<tr>
<td>How often you get fevers or wake up with wet bedsheets from sweating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>0.000</td>
<td>-</td>
<td>0.000</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sometimes</td>
<td>-0.177</td>
<td>0.501</td>
<td>-0.259</td>
<td>0.294</td>
<td>-0.743 0.225</td>
</tr>
<tr>
<td>Often</td>
<td>-0.547</td>
<td>0.012</td>
<td>-0.557</td>
<td>0.005</td>
<td>-0.948 -0.165</td>
</tr>
<tr>
<td>Always</td>
<td>-0.962</td>
<td>0.000</td>
<td>-0.915</td>
<td>0.000</td>
<td>-1.298 -0.532</td>
</tr>
<tr>
<td>How often treatment(s) get in the way of things you like to do</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>0.000</td>
<td>-</td>
<td>0.000</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sometimes</td>
<td>0.251</td>
<td>0.211</td>
<td>0.000</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Often</td>
<td>-0.121</td>
<td>0.476</td>
<td>-0.265</td>
<td>0.061</td>
<td>-0.542 0.012</td>
</tr>
<tr>
<td>Always</td>
<td>-0.715</td>
<td>0.001</td>
<td>-0.868</td>
<td>0.000</td>
<td>-1.180 -0.556</td>
</tr>
<tr>
<td>How much treatment affects your ability to attend and keep with school/university/work</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>0.000</td>
<td>-</td>
<td>0.000</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sometimes</td>
<td>0.008</td>
<td>0.976</td>
<td>0.000</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Often</td>
<td>-0.212</td>
<td>0.328</td>
<td>-0.270</td>
<td>0.102</td>
<td>-0.593 0.054</td>
</tr>
<tr>
<td>Always</td>
<td>-0.624</td>
<td>0.006</td>
<td>-0.636</td>
<td>0.000</td>
<td>-0.981 -0.292</td>
</tr>
<tr>
<td>How much of a problem you have with constipation, tummy bloating or passing wind (farting)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No-minimal prob</td>
<td>0.000</td>
<td>-</td>
<td>0.000</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Mild problem</td>
<td>-0.104</td>
<td>0.506</td>
<td>-0.027</td>
<td>0.857</td>
<td>-0.323 0.269</td>
</tr>
<tr>
<td>Moderate problem</td>
<td>-0.453</td>
<td>0.010</td>
<td>-0.396</td>
<td>0.016</td>
<td>-0.719 -0.073</td>
</tr>
<tr>
<td>Severe problem</td>
<td>-0.983</td>
<td>0.000</td>
<td>-0.976</td>
<td>0.000</td>
<td>-1.425 -0.527</td>
</tr>
</tbody>
</table>

Italicised values represent levels that have been combined to ensure monotonic ordering of coefficients.
3.2 Results of the main effect model: carers

Among carers, the coefficients for the attribute levels were combined to some extent for eight of the attributes (Table 3). The five most influential attributes on the preferred health states were (1) difficult/painful breathing, (2) pain unrelated to breathing, (3) avoidance of gastrointestinal problems, (4) feeling sad or worried and (5) fevers/night sweats. Mucus production/swallowing was important only if it occurred ‘a lot’ or in ‘huge amounts,’ and lack of energy or appetite and treatment getting in the way of ‘things you like to do’ was important if it was ‘always’ an issue. Being tired/lacking in energy and mucus production had the least influence on carer preferences overall.
Table 3: Main effects model: carers

3.3 Coefficients for people with CF versus carers

A comparison of the coefficients for the main effects models for people with CF versus carers is presented in Figure 1.
3.4 **Weights for outcomes included in the patient and proxy carer-reported instruments**

Weights for the attributes included in each instrument are presented in *Table 4*. The value for individual attribute levels can be subtracted from 100 to produce a score between 0 (worst health state) and 100 (best health state).
<table>
<thead>
<tr>
<th>Individuals $&gt;13$ years with CF</th>
<th>Carers</th>
</tr>
</thead>
<tbody>
<tr>
<td>How often it is hard or it hurts you/your child to breathe</td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>0</td>
</tr>
<tr>
<td>Sometimes</td>
<td>-9</td>
</tr>
<tr>
<td>Often</td>
<td>-11</td>
</tr>
<tr>
<td>Always</td>
<td>-21</td>
</tr>
<tr>
<td>How much mucus you/your child coughs up/swallows</td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>0</td>
</tr>
<tr>
<td>A little</td>
<td>0</td>
</tr>
<tr>
<td>A lot</td>
<td>-3</td>
</tr>
<tr>
<td>Huge amounts</td>
<td>-9</td>
</tr>
<tr>
<td>How often you/your child feels tired or lacking in energy</td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>0</td>
</tr>
<tr>
<td>Sometimes</td>
<td>0</td>
</tr>
<tr>
<td>Often</td>
<td>-1</td>
</tr>
<tr>
<td>Always</td>
<td>-4</td>
</tr>
<tr>
<td>How often you/your child doesn't feel like eating at meal times</td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>0</td>
</tr>
<tr>
<td>Sometimes</td>
<td>0</td>
</tr>
<tr>
<td>Often</td>
<td>0</td>
</tr>
<tr>
<td>Always</td>
<td>-5</td>
</tr>
<tr>
<td>How often you/your child feels pain or discomfort not related to breathing</td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>0</td>
</tr>
<tr>
<td>Sometimes</td>
<td>0</td>
</tr>
<tr>
<td>Often</td>
<td>0</td>
</tr>
<tr>
<td>Always</td>
<td>-7</td>
</tr>
<tr>
<td>How often you feel overwhelmed/your child feels sad or worried</td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>0</td>
</tr>
<tr>
<td>Sometimes</td>
<td>0</td>
</tr>
<tr>
<td>Often</td>
<td>-3</td>
</tr>
<tr>
<td>Always</td>
<td>-6</td>
</tr>
<tr>
<td>How often your child gets fevers or wakes up with wet bed sheets from sweating</td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>0</td>
</tr>
<tr>
<td>Sometimes</td>
<td>-4</td>
</tr>
<tr>
<td>Often</td>
<td>-8</td>
</tr>
<tr>
<td>Always</td>
<td>-13</td>
</tr>
<tr>
<td>How often treatment(s) get in the way of things your child likes to do</td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>0</td>
</tr>
<tr>
<td>Sometimes</td>
<td>0</td>
</tr>
<tr>
<td>Often</td>
<td>-4</td>
</tr>
<tr>
<td>Always</td>
<td>-12</td>
</tr>
<tr>
<td>How much you/your child misses school/daycare/play activities due to treatment(s)</td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>0</td>
</tr>
<tr>
<td>Sometimes</td>
<td>0</td>
</tr>
<tr>
<td>Often</td>
<td>-4</td>
</tr>
<tr>
<td>Always</td>
<td>-9</td>
</tr>
<tr>
<td>How much passing hard poos or wind (farting) is a problem for you/your child</td>
<td></td>
</tr>
<tr>
<td>No/minimal problem</td>
<td>0</td>
</tr>
<tr>
<td>Mild problem</td>
<td>0</td>
</tr>
<tr>
<td>Moderate problem</td>
<td>-6</td>
</tr>
<tr>
<td>Severe problem</td>
<td>-14</td>
</tr>
</tbody>
</table>

Table 4  Negative utility weights for the patient and carer-reported outcome instrument

3.4 Participant feedback

Of the 206 participants who completed the survey, 34 people with CF (of 97, 35.1%) and 41 carers (of 104, 39.4%) provided feedback (S8). Twenty-five people with CF and 28 carers cited difficulties in completing the survey, including confusion about the task or purpose, the complexity of the task
(n=31, 15%), scenarios that contained unlikely combinations of attributes (n=9, 4.3%) or which were outside their lived experience (n=4, 1.9%), and subject matter that was emotionally confronting (n=5, 2.4%). Six participants reported it would have been helpful to specify the duration of treatment for the pulmonary exacerbation for each question, as this would have impacted their preference selection (n=6, 2.9%).

4.0 Discussion

We report the results of a DCE quantifying the influence of distinct outcomes resulting from treatment of pulmonary exacerbations of CF on overall preferred health states from the perspective of people with CF and carers of children with CF. To our knowledge, this is the first study to attempt to quantify the relative importance of outcomes relating to treatment for these episodes and how these attributes influence decision-making.

The health outcome preferences reported here should help guide clinical decision making when agreeing shared goals of therapy for treatment of pulmonary exacerbations of CF. Difficult/painful breathing had the greatest influence on the preferred health state for people with CF and carers alike. Avoidance of gastrointestinal problems were also strongly influential.

There were some differences in the prioritisation of attributes by people with CF and carers, consistent with the patterns reported for other chronic diseases (12, 13). The impact of treatment getting in the way of ‘things you like to do’ had a stronger influence on preferred health states for people with CF compared to carers, whose preferences were impacted more by the presence of pain unrelated to breathing, and anxiety/worry experienced by children in their care. The DCE did not define the origins of pain ‘unrelated to breathing’; consequently, participants may have included pain related to medical procedures such as intravenous cannulation. These findings require further investigation. While not examined in this study, the perspectives of people with CF and carers may
differ from those of health professionals. For example, coughing up/swallowing mucus was perceived negatively by people with CF and carers. This may be counterintuitive from the perspective of health professionals, given that the clearance of mucus is considered desirable to avoid stasis of airways secretions and concomitant infection.

The attributes included in the instruments developed in this study did not consistently match items included in alternative patient-reported outcome measures currently available. These include the cystic fibrosis questionnaire-revised (CFQ-R) (14, 15); the version for people ≥14 years old evaluates 50 items over 9 domains capturing symptoms and function (for the previous 2 weeks), and the CRISS-CFRSD; this evaluates 16 symptoms over the past 24 hours. Sleep, exercise tolerance and social isolation and cough, wheezing, frustration, time spent lying/sitting down and chest tightness are included in the CFQ-R and CRISS-CFRSD (15, 16), respectively, but not in the instruments presented here. Outcomes included in the instruments developed in this study were chosen based on those that were prioritised as important to people affected by disease and which captured separate rather than common aspects of the underlying pathophysiological disease process. The rationale for this was to minimise the burden on participants by achieving a parsimonious instrument, avoiding redundancy in attributes that capture similar outcomes without contributing additional information about the overall patient experience.

Several limitations of our study must be acknowledged. First, most participants were female and resided in Australia, which may limit the generalisability of results. Second, the perspectives of young children were not directly solicited, and while adolescents with CF were invited, few participated. While there is some data to suggest that the perspectives of carers align with those of the children under their care for other diseases (17, 18), these perspectives may not always concord. Third, selection bias may have occurred for two reasons; (i) only people with access to the internet were able to participate, and (ii) remuneration was not offered when the survey first commenced
and was not advertised; this may have been a disincentive to participate. Fourth, although the survey was developed with input from people affected by CF, a few participants reported that some combinations of attribute levels were unlikely or implausible. Finally, survey completion rates were not optimal. While the specific reasons for this are not known, the feedback provided by those who did finish the survey suggests that the task was cognitively challenging; this was likely to be a major contributing factor to drop-out.

This study quantifies the relative influence of distinct outcomes on the overall preferred health state resulting from treatment of pulmonary exacerbations from the perspective of people with CF and carers. These data should inform clinical decision making. It is intended that the patient and proxy carer-reported weighted instruments developed here will be used in subsequent trials of pulmonary exacerbations of CF, however these will require subsequent validation. Future studies should focus on obtaining preference information from children, including adolescents, and exploring how these preferences compare to those of their carer(s).

Acknowledgements
The authors would like to acknowledge the individuals who contributed their time and subject matter expertise to this study, including development of the DCE, particularly people with lived experience of the disease. This work was supported by Cystic Fibrosis Australia and the Perth Children’s Hospital Foundation (New Investigator grant; 2017). Fellowship support was provided by the NHMRC (GNT1150996), MRFF (APP1193796), the Wesfarmers Centre and the Perth Children’s Hospital Foundation. Chris Blyth is also supported by an NHMRC EL2 Fellowship grant (APP1173163).

Funding
This research was supported by a Perth Children’s Hospital Foundation New Investigator Grant (9772). CM is supported by an NHMRC post-graduate scholarship (GNT1150996), and top-up grants from the Wesfarmers Centre (top-up grant) and the Perth Children’s Hospital Foundation (9757). TS is supported by a Career Development Fellowship from the National Health and Medical Research Council (GNT1111657). CB is also supported by the National Health and Medical Research Council (1111596/1173163).

Conflicts of interest
ASm reports grants from Vertex, as well as speaker honoraria and expenses from Vertex and Teva, outside the submitted work. In addition, ASm has a patent issued “Alkyl quinolones as biomarkers of Pseudomonas aeruginosa infection and uses thereof”. Other authors have no conflicts of interest to declare.

Author contributions
TS, RN and CM were responsible for the overall DCE study concept. TS, CM, RN, AS, SM and SW elaborated the study protocol for the DCE. CM drafted the manuscript. All authors revised and approved the final manuscript.

Ethics approval
Ethics approval for all aspects of this study was granted by the Western Australia Child and Adolescent Health Service Human Research Ethics Committee [RGS903].
References


Chapter 9: Conclusions and directions for future research

9.1 Chapter overview

In this final chapter, I summarise the main findings from this thesis and outline recommendations for future research.

9.2 Summary of the main findings, strengths and limitations and conclusions from this thesis

9.2.1 Generation of a clinical researcher’s guide for applying an estimands framework when designing clinical trials

Use of the estimands framework is encouraged to help ensure the design and statistical analysis of clinical trials aligns with their objective, and thereby generate evidence to reliably inform clinical decision making. This framework has been used to inform the trial design for BEAT CF. While use of this framework has been promoted in the statistical literature via a recent addendum to the International Council for Harmonisation Guidance E9(R1) on the Statistical Principles for Clinical Trials (ICH, 2019), at the time of writing this thesis the framework had little recognition among many clinical researchers. This manuscript introduces the concept of an estimand for clinical researchers and describes the five attributes of an estimand, using hypothetical examples from trials of CF. These attributes include the population of interest, the endpoint, a description of the treatment intervention, intercurrent events, and the population level summary. I further differentiate between an outcome, the effect of interest but which may not be directly measurable, and an endpoint, how the outcome is captured, albeit perhaps imperfectly. I also distinguish these concepts from the population level summary of the treatment effect which is how the causal effect of an intervention is measured. Despite previous attempts to formalize the distinct definitions of these concepts, they are still often used interchangeably in the clinical research literature. This guide is intended to explain their difference, and provide a rationale and describe the application of the estimand framework among clinical researchers, who are encouraged to use it when designing trials in order to improve the transparency and the value of their research (Akacha, 2017).
9.2.2 Summary of the characteristics, properties and evolution of endpoints used in late phase clinical trials

Selecting outcomes and their corresponding endpoints to evaluate treatment effects in clinical trials is a critical component of clinical trial design. Firstly, the selected outcomes should be meaningful to people affected by the condition and those involved in practice and policy to ensure the applicability of the research to decision-making. The choice of endpoints based on these outcomes may be improved by understanding their characteristics and properties. This review aims to guide those responsible for designing trials, or for interpreting and translating the evidence generated from trials into practice and policy.

9.2.3 Systematic review of the outcomes and endpoints reported in pulmonary exacerbation trials in people with CF

This is a comprehensive review of the outcomes and their corresponding endpoints which have been reported in pulmonary exacerbation trials in people with CF. I build on previous distinction between outcomes and endpoints I introduced in the earlier chapters in order to group endpoints by the outcome they attempt to capture. This review highlights outcomes that are most likely to be meaningful; that is, those that reflect, prima facie, how a person feels, functions or survives. There were 144 studies that met inclusion criteria. A wide range of outcomes and corresponding endpoints are reported. Death, QoL and many patient-reported outcomes are likely to be meaningful. The predominant endpoint used to capture lung function, the forced expiratory volume in one-second (FEV₁), is a validated surrogate that is correlated with risk of death and quality of life. The extent of structural lung disease was also found to correlate with lung function, pulmonary exacerbations and risk of death. There was no evidence found of a correlation between airway microbiology or biomarkers with clinically meaningful outcomes; the value of these surrogates for decision-making is consequently unproven.

One important finding of this review was the lack of validated endpoints for use in children less than six years old. For example, lung function measured as the FEV₁ (or volume of air that can be forcibly exhaled in one second) is the most common outcome evaluated in trials for people with CF (McLeod, 2020); however, children under 6 years old generally have well preserved lung function and can rarely perform spirometry, so this endpoint is irrelevant for them. The absence of an agreed outcome in this age group is a barrier to clinical studies in this group.
This systematic review is an important step in the development of a core outcome set (COS), a standardised collection of outcomes that should be reported in all trials for pulmonary exacerbations in people with CF, including young children. Development of a COS with key stakeholders will be a focus for post-doctoral studies (see section 9.3.1).

9.2.4 Systematic review of the measurement properties of tests and tools used in trials of CF

This is the first systematic review of the measurement properties of tests and tools used to capture outcomes in studies of pulmonary exacerbations in children and adults with CF. The reliability, validity and responsiveness of these tools was assessed. There were 118 studies evaluating the measurement properties of 74 tests and tools identified. Tests or tools were categorised as patient-reported outcome measures (PROMs) capturing QoL or other patient-reported outcomes, clinical scores, radiological scores, or tests capturing lung function, functional exercise performance, CF transmembrane regulator (CFTR) function, or sputum characteristics.

A COS, that is a consensus set of tests and tools for measurement of outcomes in trials in people with CF is recommended; this should be developed with key stakeholders including people living with disease. This would be likely to improve the consistency of reporting and measurement of similar outcomes, allowing comparison and synthesis of evidence across studies and improving the value of the research that is conducted. This will be a focus for post-doctoral studies (see section 9.3.1).

9.2.5 Development of a novel method to select meaningful outcomes for evaluation in clinical trials

I propose a novel method for the selection of outcomes and endpoints used to assess treatment effects in clinical trials. The approach is rooted in causal inference. This method involves choosing outcomes and their corresponding endpoints that: (i) are identified as meaningful to key stakeholders including those with lived experience of the disease, (ii) capture an aspect of the underlying pathophysiological process (if multiple outcomes are studied, ideally these should capture separate rather than shared aspects of the pathophysiological processes) and (iii) are likely to be impacted by the intervention under study.

The rationale for this approach builds on Prentice’s argument (Prentice, 1989) that the evaluation of
outcomes and endpoints that are not causally related to the disease of interest and dependent on the intervention being studied may produce misleading information. This method is likely to be helpful for informing the selection of outcomes in other health domains and may improve the value of the research that is conducted.

9.2.6 Identification of preferred health outcome states resulting from treatment of pulmonary exacerbation episodes and the development of weighted patient and proxy carer-reported outcome measures capturing symptoms and functional impacts in children and adults with CF

The is the first DCE experiment regarding treatment for pulmonary exacerbations to be conducted in people affected by CF. Overall, 362 people participated in the DCE (167 people with CF and 195 carers), of whom only 206 completed the survey (57%). Most respondents were female and resided in Australia. Difficult/painful breathing had the greatest impact overall on the preferred health state among people with CF and carers alike. The desire to avoid gastrointestinal symptoms also appeared to heavily influence the preferences of participants.

There are two important implications of this study. First, the health outcome preferences of participants captured in this study should guide clinical decision making for the treatment of pulmonary exacerbation episodes, and therefore also guide the selection of outcomes for clinical trials of pulmonary exacerbations of CF.

Second, this study has led to development of a MAOI; one version of the MAOI is based on patient-reported outcome measures for use in adults and adolescents, and another is based on proxy carer-reported outcome measures for use in children. The ten attributes included in each instrument have been identified as important to people affected by CF and aim to broadly capture impacts on symptoms and functional capacity in the context of treatment for pulmonary exacerbation episodes. I plan to formally assess the validity of these instruments in BEAT CF prior to their broader use.

9.2.7 Capacity building and protocol for the development of a COS for infective exacerbation trials in CF

In addition to the outputs described above, this research has advanced the field of CF research by strengthening stakeholder engagement by helping to establish an international research network for the development of COS for CF trials. This was possible due to collaborations that were built during this PhD. Specifically, I formed partnerships with experts in the field of CF (including Professor
Alan Smyth, United Kingdom) and COS development (including Professor Allison Tong, University of Sydney, Australia). The international network’s first objective is to develop an international COS to define best-practice for inclusion of outcomes in pulmonary exacerbation trials. This should improve the consistency and transparency in the selection and reporting of meaningful outcomes in pulmonary exacerbation trials, and improve the robustness of future clinical trials, which will lead to improved outcomes for young children living with CF. Most crucially, the COS will be adopted into the ongoing BEAT CF platform trial, allowing for children <6 years old to participate for the first time.

9.3 Where does this thesis sit with relation to the science of clinical trials?

Randomised controlled trials are rightfully considered the gold standard approach for generating evidence designed to inform clinical practice and policy. However, randomisation alone doesn’t guarantee the applicability of trial results to guide clinical decision-making. The reasons for this are multiple and include (i) failure to align the selection of outcomes and endpoints, and the statistical approach (including the handling of post-randomisation events) with the objectives of the trial, namely to inform a treatment decision or policy (Akacha, 2017), (ii) the selection of outcomes and endpoints of little meaning to people with lived experience of disease or other relevant stakeholders (Treweek, 2009), (ii) the use of explanatory designs focused on exploring mechanistic outcomes, over pragmatic designs which focus on patient-relevant treatment outcomes observed across broad ‘real world’ clinical contexts which are necessary to inform decision-making (Norton, 2021; Zwarenstein, 2009; Chalkidou, 2012).

While the needs for trials that are pragmatic and which involves consumers in their design has been recognised and advocated for a number of years, this thesis builds upon this principle by proposing a formal approach to trial design that is deeply embedded in causal reasoning and inference. In this sense, it is a departure from traditional empiricist perspective on clinical trials which focus on treatment and endpoint, ignoring the (pathophysiological) mechanisms that underlie the mediation and capture of patient-relevant treatment outcomes. While mechanisms should not be the target for the measurement of treatment effects, and explicit consideration of how treatments affect various outcomes and how those outcomes are captured as endpoints, might help to improve trial design, in particular when it comes to selecting outcomes and endpoints. I have argued that this approach is valuable for generating evidence to inform decision-making in clinical practice and policy, using trials of pulmonary exacerbations in people with CF as an exemplar. Specifically, I have recommended (i) using a estimands framework to clarify the research objective, and to ensure the
design of the trial, including the statistical methods, align with that objective and that the results can therefore be translated to inform clinical decision-making and (ii) selecting outcomes and their corresponding endpoints that relate to causal pathways of disease which are impacted by the intervention being studied, and which are recognized as meaningful to those with lived experience of the disease.

This approach is likely to improve the value of the research that is conducted when applied across any research domain and ultimately improve outcomes for people living with the disease/condition being studied.

### 9.4 Directions for future research

Future research is required to establish a COS to inform the design of all studies of pulmonary exacerbations in people with CF based on the shared priorities of key stakeholders. To conclude this thesis, a protocol for establishing such a COS for trials of pulmonary exacerbations in people with CF has been developed and included. Funding has been secured to support this initiative.

*Submitted journal article*

A protocol for establishing a core outcome set for evaluation in trials of pulmonary exacerbations in people with cystic fibrosis

Charlie McLeod¹-³, Alan Smyth⁴, Mitch Messer¹, Martin Howell⁵, Andre Schultz⁶,⁷, Jamie Wood⁸, Richard Norman⁹, Christopher Blyth¹-³,¹⁰, Steve Webb¹¹,¹², Tom Snelling⁵,¹³, Allison Tong¹²,¹⁴

¹Wesfarmers Centre for Vaccines and Infectious Diseases, Telethon Kids Institute, University of Western Australia, Nedlands, WA, Australia

²School of Medicine, University of Western Australia, Nedlands, WA, Australia

³Infectious Diseases Department, Perth Children’s Hospital, Nedlands, WA, Australia

⁴Evidence Based Child Health Group, University of Nottingham, Queens Medical Centre, Nottingham, UK

⁵Sydney School of Public Health, The University of Sydney, Sydney, Australia

⁶Wal-yan Respiratory Research Centre, Telethon Kids Institute, University of Western Australia, Nedlands, Australia

⁷Department of Respiratory and Sleep Medicine, Perth Children’s Hospital, Nedlands, Australia

⁸Abilities Research Center, Department of Rehabilitation and Human Performance, Icahn School of Medicine at Mount Sinai, New York, United States of America

⁹School of Public Health, Curtin University, Bentley, Australia

¹⁰Pathwest Laboratory Medicine, QEII Medical Centre, Nedlands, Australia

¹¹St John of God Hospital, Subiaco, Australia

¹²School of Population Health and Preventive Medicine, Monash University, Melbourne, Australia
13Menzies School of Health Research, Casuarina, Australia

14Centre for Kidney Research, The Children’s Hospital at Westmead, Westmead, Australia

**Corresponding Author:**

Charlie McLeod; Perth Children’s Hospital, 15 Hospital Avenue, Nedlands, WA, Australia;

E: charlie.mcleod@health.wa.gov.au; P: +61 8 6456 2222
Abstract

Background

Pulmonary exacerbations are associated with increased morbidity and mortality in people with cystic fibrosis (CF). There is no consensus about which outcomes should be evaluated in trials of pulmonary exacerbations or how these outcomes should be measured. Outcomes of importance to people with lived experience of the disease are frequently omitted or inconsistently reported in trials, which limits the value of such trials for informing practice and policy. To better standardise outcome reporting and measurement, we aim to develop a core outcome set for trials of pulmonary exacerbations in people with CF (COS-PEX) and consensus recommendations for measurement of core outcomes.

Methods

Preliminary work for development of COS-PEX has been reported, including (i) a systematic review of outcomes reported in existing trials of pulmonary exacerbations (ii) workshops with adults, adolescents and carers of children with CF within Australia and (iii) a bayesian knowledge expert elicitation workshop with health professionals to ascertain outcomes of importance. Here we describe a protocol for the additional stages required for COS-PEX development and consensus methods for measurement of core outcomes. These include; (i) an international two-round online Delphi survey involving key stakeholders (ii) consensus workshops to review and endorse the proposed COS-PEX and (iii) consensus workshops to agree methods for measurement.

Discussion

COS-PEX should improve the consistency and transparency of outcome reporting for pulmonary exacerbation trials in CF, facilitate comparison of data across studies, and thereby improve the quality and value of research in this field for informed decision-making.
Keywords:

Cystic fibrosis, core outcome set, patient-centred outcomes, outcomes research
Background
Cystic fibrosis (CF) is a life limiting disease characterised by episodic pulmonary exacerbations which are thought to drive progressive lung damage (1). Treatment for these episodes is complex and generally involves a combination of antimicrobials, therapies to improve airway clearance (including chest physiotherapy and muco-active agents), optimisation of nutrition, and possibly anti-inflammatories (1-5). Of the ten Cochrane reviews evaluating trials of treatment strategies for pulmonary exacerbations, most were inconclusive and many controversies about treatment remain (1, 2, 4, 6, 7). Synthesis of data is impeded by inconsistency in the selection of outcomes in these studies and how they have been assessed, and by the reporting of outcomes that might not be meaningful to people living with disease.

A core outcome set (COS) represents a minimum set of agreed outcomes derived from broad stakeholder consensus for measurement and reporting in all trials for a specific condition (8). Trialists can add other outcomes relevant to the trial. The Core Outcome Measures in Effectiveness Trials (COMET) initiative was established in 2010 to improve consistency in the selection of meaningful outcomes when designing studies, to facilitate collaboration, avoid duplication, and to improve the value of the research that is conducted (9).

We have reviewed the range of outcomes previously reported in studies of pulmonary exacerbations in CF in people with CF (10) and the methods used for measuring these outcomes [McLeod; accepted for publication]. We have engaged Australian stakeholders including clinicians, people 13 years and above with CF and carers of children less than 18 years of age with CF to identify outcomes of importance to them. The top ten outcomes capturing symptoms or functional capacity from the perspective of people affected by CF
were difficulty/painful breathing, sputum production and clearance, fatigue, appetite, pain, motivation/demoralisation, fevers/night sweats, treatment burden, inability to meet personal, school, or work goals and avoidance of gastrointestinal symptoms (constipation, bloating and flatulence) [McLeod, accepted for publication]. Our group has also conducted a bayesian expert knowledge elicitation workshop to elicit outcomes and of importance to health care professionals [Snelling, unpub]. It is unknown however if these priorities are shared by stakeholders outside Australia.

We have therefore established an international steering committee to oversee the development of COS. The steering committee comprises a range of subject matter experts from different countries including people with CF from diverse backgrounds, health care professionals and health care commissioners, researchers, people affected by disease, and people involved in the dissemination and translation of research findings into practice and policy. Development of a core outcome set for trials of pulmonary exacerbations of CF (COS-PEX) will be the first initiative to be driven by this group. Our aim is to establish a COS to inform the design of all studies of pulmonary exacerbations in people with CF based on the shared priorities of key stakeholders. It is expected that this research will aid the selection of meaningful outcomes for evaluation in relevant studies in order to optimise the value of the research and to minimise research waste (11).

**Method**

COS-PEX is registered in the COMET database (12). The protocol for generation of this COS has been adapted from the Core Outcome Measures in Effectiveness Trials (COMET) framework (12). In addition to the work described above, development of the COS-PEX and consensus methods for measurement of core outcomes will involve three additional steps;
these are presented in *Figure 1*. Ethics approval has been provided by the Child Adolescent Health Service (RGS903). The project will be conducted in accordance with the Core Outcome Set Standards for Development (COS-STAD) (13). Results will be reported according to the Core Outcome Set Standards for Reporting (COS-STAR) (14).

**Figure 1**  
Method for development of COS-PEX

**Step one**: Online Delphi surveys

Outcomes will be defined according to the taxonomy proposed by COMET (15). An outcome will be defined as a measurement or observation used to capture and assess the effect of treatment such as an assessment of side effects (risk) or effectiveness (benefits) (16). Outcome domains will be defined as an aspect of health that is likely to be impacted by a health care intervention (15). Outcomes identified from the systematic review (10) and
preliminary workshops [McLeod, unpub] will be mapped to domains by members of the steering committee in order to structure a list for evaluation in the eDelphi.

An international two-round online Delphi survey (eDelphi) will be conducted to generate consensus about outcome domains of importance to key stakeholders (see Figure 1). This will involve two rounds of 20-30 minute surveys answered anonymously; participants will be asked to rate approximately 30 outcomes randomised by block allocation according to their perceived importance. There will be two versions of the survey, including (i) for children with CF aged between eight and 18 years of age (ii) for adults with CF and other stakeholders. Each version will evaluate identical outcomes, although the wording will be tailored to the targeted population’s age and role, and the paediatric version will include a picture illustrating each outcome. Both versions will include plain language definitions for the outcomes presented. The version for children will be pitched at grade five reading level according to the Flesch-Kincaid Index (10 years of age). Both versions will be piloted to a minimum of five people and feedback will be incorporated prior to finalisation.

The eDelphi method has been validated as a reliable approach for achieving consensus on core outcome sets for various health conditions (17). This method involves participants contributing subject matter knowledge independently, and then having the opportunity to revise their responses based on the feedback and opinions offered by other respondents. Participants do not interact directly with each other, thereby avoiding domination of the discussion by few contributors. Data will be reported according to the checklist recommended by Sinha et al (18); this will include a discussion of the size and composition of the panel, the Delphi method and the results.
Participants and recruitment

Subject matter experts including people with CF from diverse backgrounds, carers of children less than 18 years of age, health care professionals, researchers, journal editors, policy makers and regulatory and pharmaceutical authority representatives will be eligible to participate. We will employ multiple recruitment strategies to ensure inclusivity and diversity of representation, including recruitment though (i) participating medical facilities, (ii) consumer and patient networks, including, but not limited to, advertising via email and social media including Facebook and Twitter, (iii) recruitment by investigators with lived experience of disease. We will also use snowballing strategies enabling participants to extend an invitation to other relevant stakeholders to participate. Individuals will also be able to access the survey directly at https://www.beatcf.org.au. Monetary remuneration may be offered to participants as compensation for their time to promote the representativeness of the sample.

There are no recommendations available to guide the determination of a minimal sample size for Delphi surveys for the purposes of developing COSs (19). Our target sample size will be a minimum of 250 respondents, which is just above the lower participant limit used to develop COS reported in the literature; we will aim to include at least 200 people with CF, 25 carers of children with CF, and 25 health professionals.

Survey administration

A link will be provided to access the online survey. The surveys will be developed in research software, Qualtrics (20). Responses will be anonymous and all data will be non-identifiable. Participant information and consent will be included online. Individual consent will be required for participation, and consent from guardians of children between eight and
18 years old will also be required for children to participate. Individuals will be able to exit the online surveys at any time prior to submission of their responses. After this time, anonymisation will make it impossible to withdraw their responses.

Data collection – round one

Each participant will rate the outcome domains according to the nine-point Likert scale suggested by GRADE (21); ratings from one to three reflect outcomes of “limited importance”, ratings from four to six include outcomes that are “important but not critical”, and ratings between seven to nine indicate outcomes of “critical importance”. An “unable to score” option will also be available. A best-worst evaluation exercise will also be included in the survey; this is an established method that can be used to calculate the relative importance of juxtaposed outcomes (22). Participants will be given the opportunity to provide feedback including a rationale for their answers and to suggest additional outcome domains of importance not included in the round one survey.

Basic demographic information including sex, age and stakeholder group will be requested. Clinicians will be asked how many full-time equivalent years of clinical experience they have had in caring for people with CF, and people with CF will be asked several questions to help categorise their severity of disease (such as number of hospitalisations over the past 12-months and lung transplant status).

Responses will be considered separately for the following groups: (i) children and their carers and (ii) adults with CF and other stakeholders. Outcomes with a mean and median score greater than 7 based on responses from at least 70% of respondents will be included in round
two, as well as new outcome domains that are suggested by more than ten per cent of participants.

*Data collection – round two*

In the round two survey, participants will be presented with a graph of the distribution of scores for each outcome for the respective groups. An explanation to aid interpretation of the graph will be provided, including an animated explanation for the paediatric survey. Comments made by participants in the round one survey will also be supplied, and the individuals’ own responses will be highlighted. Participants will be asked to repeat the rating exercise using the same method described for round one. A best-worst exercise will also be repeated.

Prior to completion of the survey, participants will be invited to register their interest in participating in the consensus workshop(s) described in *step two* by supplying their name and email address.

*Data analysis*

Quantitative analysis will be conducted using STATA V.13. The analysis will involve calculating the distribution of scores and the mean, median and proportion of scores ranked for each Likert category, and the overall ranking of the outcome domains according to the responses from the two groups. Results for the paediatric and adult survey, as well as the differences in responses between people with CF compared to other stakeholders will be compared. The criteria for inclusion in each COS is based on the recommendations specified by OMERACT (23). Domains with a median and mean of more than seven based on responses from 70% or more from people with CF/carers and health professionals/policy
makers for the respective surveys will be included as “middle tier” outcomes at a minimum; the top three to five core outcome domains will be selected based on means, medians and proportions (see Figure 2). If the thresholds for inclusion are modified post hoc, these will be reported to ensure transparency. The utility function of outcomes examined in the best-worst exercise will be calculated for all outcomes using conditional logit regression analyses.

**Figure 2**  
*COS-PoC*

**Step two: Consensus stakeholder workshops**

Two consensus workshops chaired by members of the steering committee will be conducted to review the proposed COS based on the results obtained from the two-stage eDelphi. The first workshop will focus on outcomes for adults with CF and the second workshop will focus on outcomes for children with CF. The workshops will be up to two hours in duration and will occur via videoconference, owing to infection control restrictions which preclude mixing of people with CF. All attendees will be invited to participate as investigators rather than research subjects; consent will therefore not be required.

**Participants and recruitment**

Recruitment methods will be the same as those employed for the eDelphi. Individuals who register their interest in attending will be provided with a copy of the written results of the
relevant eDelphi. Those attending the paediatric COS workshop will also be provided with a link to an online animation to explain the results.

Data collection
Run sheets will be developed for the COS-PEX workshops. An assigned member of the investigator group will record notes on the group dynamics and interaction between participants. The anticipated workshop format will involve (i) welcome and presentation of results from step one, (ii) breakout discussion in groups comprising approximately ten participants facilitated by a moderator to discuss the differences in results between groups and any identified issues, to resolve any uncertainties, and to agree on the proposed scope of the COS (including the specific CF population(s), the setting and the type of intervention(s) for which the COS is likely to be relevant), (iii) a summary of each group discussion will be reported back to the larger group and (iv) participants will be asked to endorse the final agreed COS. All participants will be given the opportunity to ask questions and discuss any differences or similarities in opinion.

Data analysis and reporting
Transcripts of the breakout discussions will be entered into the HyperRESEARCH software. CM will code this data and use thematic analysis to explore the range of perspectives for core outcome domains and will report key recommendations and anticipated challenges for implementation for the COS.

A preliminary plain language report will be disseminated to all workshop participants and relevant stakeholder groups for the COS-PEX and posted via the https://www.beatcf.org.au website to invite public comment for a period of two weeks. The steering committee will then
finalise and endorse the final COS prior to dissemination of the results by peer-reviewed publication.

**Step three: Consensus methods for measurement of the core outcome domains**

Up to three online workshops will be convened by members of the steering committee to develop consensus recommendations regarding the measurement of the core outcomes identified for each COS based on the methods previously identified from a systematic review of tests and tools used to measure outcomes in trials of CF [McLeod, accepted for publication]. All attendees will be invited to participate as investigators; consent will therefore not be required.

**Participants and recruitment**

Recruitment methods will be the same as those employed for steps one and two.

**Data collection**

The anticipated workshop format will involve (i) presentation of available tests and tools for measurement of the core outcomes included in each COS identified by a recent systematic review and feasibility considerations, (ii) breakout discussion in groups comprising approximately ten participants facilitated by a moderator to discuss the utility of different tests and tools for measurement and to agree on the favoured tests or tools and the scope for use (including the specific CF population(s) and setting), (iii) a summary of each group discussion will be reported back to the larger group and (iv) participants will be asked to endorse the final agreed tests of tools for measurement. All participants will be given the opportunity to ask questions and discuss any differences or similarities in opinion.
Data analysis and reporting

Workshops will be audio-recorded. Data will be imported to the HyperRESEARCH software. These data will be qualitatively evaluated using thematic analysis to identify themes and recommendations. These results will be discussed among members of the steering committee.

A preliminary report will be drafted and disseminated to all workshop participants and relevant stakeholders and posted via the https://www.beatcf.org.au website to invite public comment for two weeks. The steering committee will review and endorse the final recommendations regarding consensus methods for measurement of core outcomes prior to dissemination of the results via peer-reviewed publication.

Discussion

While a COS has been established for a number of other chronic conditions including rheumatological (24) and kidney disease (25, 26), this is the first international collaboration to aim to establish a COS for evaluation in pulmonary exacerbation trials in people with CF and consensus recommendations for the measurement of core outcomes.

COS-PEX is the first of several proposed COS initiatives planned by the international steering group established to oversee this project. Members of the steering committee include a range of diverse representatives from around the world who are recognised contributors to the field of CF research and who are situated to promote implementation of COS-PEX in future studies.
Development of COS-PEX is expected to improve the value of research in this field and minimise waste by ensuring that the outcomes evaluated and reported are meaningful to all relevant stakeholders, and most importantly people with CF. It is also expected that this work will improve the transparency of outcome reporting and assessment of studies, optimise research engagement of people affected by disease, improve the acceptance and translation of research findings, and help facilitate comparison of data and synthesis of evidence across studies.

Acknowledgements

CM is supported by an NHMRC post-graduate scholarship (GNT1150996), and top-up grants from the Wesfarmers Centre (top-up grant) and the Perth Children’s Hospital Foundation (9772 & 9757). TS is supported by a Career Development Fellowship from the National Health and Medical Research Council (GNT1111657). CB is also supported by the National Health and Medical Research Council (1111596/1173163).

Conflicts of Interest

The authors have no conflicts of interest to declare.
References


5. Waters V, Ratjen F. Combination antimicrobial susceptibility testing for acute exacerbations in chronic infection of Pseudomonas aeruginosa in cystic fibrosis. Cochrane Database Syst Rev. 2015(11):CD006961.

6. Elphick HE, Scott A. Single versus combination intravenous anti-pseudomonal antibiotic therapy for people with cystic fibrosis. 2016(1469-493X (Electronic)).


20. team Q. Survey tool 2021 [Available from: https://www.qualtrics.com/au/?utm_source=google&utm_medium=ppc&utm_campaign=APJ-ANZ-Brand-Qualtrics-Brand&utm_keyword=qualtrics&MatchType=e&adid=492567885792&utm_content=492567885792&adgroupid=112314781144&campaignid=10690854453&AdGroup=%7BAdGroup%7D&BidMatchType=%7BBidMatchType%7D&Target=&targetid=kwd-350281095527&Device=c&devicemodel=&loc_physical_ms=9070544&Placement=&querystring=%7Bquerystring%7D&network=g&adposition=&GCLID=EAIaIQobChMI1cCXpZC07wIVhjMqCh3UvQGHEAAYASAAEgLy9_D_BwE&gclid=EAIaIQobChMI1cCXpZC07wIVhjMqCh3UvQGHEAAYASAAEgLy9_D_BwE.


References


Amin R, Waters V. Antibiotic treatment for Stenotrophomonas maltophilia in people with cystic fibrosis. Cochrane


162


Rand S, Prasad S, Main E. New incremental field step-test (iStep) is valid and feasible in measuring near maximal exercise performance in children with cystic fibrosis. Physiotherapy. 2015;101:eS931–eS932. Available from: http://dx.doi.org/10.1016/j.physio.2015.03.1777


United States Department of Health and Human Services Food and Drug Administration. E6(R2) Good Clinical
United States Food and Drug Administration. Table of surrogate endpoints that were the basis of drug approval or licensure. 2020[27 September 2020]. Available from: www.fda.gov/drugs/development-resources/table-surrogate-endpoints-were-basis-drug-approval-licensure


Willgoss T, Trigg A, Meysner S, Kitchen H, Humphrey L, Blankenburg M. Understanding the suitability of cystic
fibrosis-specific clinical outcome assessments for clinical trials and to support medical product labelling. 2015;18(3):A286.


