Cystic fibrosis (CF) is a complex genetic disease characterized by death from loss of lung function. Therapies target pathophysiologic changes associated with pulmonary disease progression. Although therapeutic mechanisms differ, efficacy demonstration is limited to a few accepted outcome measures, each with shortcomings that are becoming more pronounced as CF population health improves. Pulmonary function improvement (as forced expiratory volume in 1 s \( \text{FEV}_1 \)) and reduction of pulmonary exacerbation risk are commonly used outcomes. Changes in \( \text{FEV}_1 \) decline rate, quality of life, linear growth and/or weight gain are less utilized outcomes. Validated outcomes tend to work best in subjects with more aggressive or advanced lung disease and less so in healthier subjects. Assays of effects on primary therapeutic targets have yet to be validated as surrogate measures of clinical efficacy. As CF population health improves, it will become increasingly difficult to employ current clinical outcome measures to demonstrate efficacy.

Keywords: biomarkers • clinical trials • cystic fibrosis • exacerbation • \( \text{FEV}_1 \) • lung disease • outcomes

Cystic fibrosis
Cystic fibrosis (CF) is a life-shortening genetic disease [1] for which >80% of mortality today can be attributed directly or indirectly to a loss of pulmonary function [2,3]. CF results from inheritance of one of the more than 1850 identified mutant alleles [201] of the \( CFTR \) gene from each parent [4], and affects more than 70,000 people worldwide [2,3,5,6].

Reduced or absent CFTR protein function dysregulates electrolyte and fluid balance across CF epithelial cells of the respiratory, gastrointestinal (GI), hepatobiliary and reproductive tracts, as well as of the pancreas and sweat ducts [7]. GI manifestations and exocrine pancreatic insufficiency combine to create substantial nutritional challenges that can largely be managed with a high fat diet, pancreatic enzyme replacement therapy and nutritional supplementation [8,9]. Airway surface liquid dehydration causes impaired mucociliary clearance (MCC) [10] and airway obstruction, predisposing individuals to complex opportunistic infection [11] and inflammation resulting in progressive, irreversible and ultimately fatal lung damage [12]. For these reasons, there has been a tremendous need for the development and testing of interventions intended to reduce or delay CF lung disease. In this review, we describe the strengths and weaknesses of clinical trial end points that are currently accepted by regulators for the demonstration of efficacy of CF pulmonary therapies, how changes in CF population demographics and treatment algorithms have created a need for additional efficacy end points that can be employed earlier in the lung disease process and how candidate efficacy end points may meet these needs.
CF clinical trial end points identified as ‘appropriate for demonstrating tangible benefit of a new therapeutic agent’ have not changed dramatically in the nearly two decades since they were enumerated at a consensus conference organized by the Cystic Fibrosis Foundation and the US NIH: stable or improved pulmonary function, decreased frequency of pulmonary exacerbations, quality of life (QoL) improvement and, for younger patients, growth improvement [13]. Three of these end points (exacerbation, QoL and growth) are considered ‘true’ clinical end points that directly measure how a patient ‘feels, functions or survives.’ The fourth, change in pulmonary function, is a ‘surrogate’ clinical end point in that an individual may not be able to perceive a modest change in their pulmonary function [14].

A distinction must be made for CF clinical trial end points between observation of statistical significance (i.e., that a difference in an outcome observed between treatment groups is highly unlikely to be the result of random chance) and clinical significance (that an observed difference in an outcome is clinically meaningful). Although there are substantial data with which to design CF clinical trials in order to assure detection of a given magnitude of treatment effect with a given end point, there is a paucity of data as to what magnitude of treatment effect is clinically meaningful for many end points. Most controlled CF trials have employed statistical tests of superiority of active treatment over a placebo, which does not require identification of a threshold difference for clinical significance. In truth, the relatively small pool of CF subjects available for clinical trial enrollment limits an investigator’s ability to power studies to detect small (i.e., potentially clinically insignificant) treatment differences. If a statistically significant difference is observed, then clinicians are left to ponder clinical significance. However, as more CF therapies become available and there is a greater need to test investigational therapies for their equivalence/non-inferiority to active (as opposed to placebo) therapies, clinically meaningful changes in end points will have to be established and agreed upon prior to study initiation.

Improvement/stabilization of pulmonary function

Pulmonary function expressed as forced expiratory volume in 1 s (FEV$_1$) and that fraction of an individual’s FEV$_1$ retained compared with a healthy reference population of similar height, age and sex (FEV$_1$% predicted) are highly (some would say disproportionately [15,16]) prominent measures in CF clinical care and research. As noted, more than four out of five CF deaths today are the direct or indirect result of loss of pulmonary function [2,3]. Perhaps not surprisingly, lower FEV$_1$% predicted has been associated with lower 2-year CF survival [17], and individuals with CF dying at younger ages have, on average, greater average lifetime rates of decline of FEV$_1$% predicted than those dying at older ages [18]. Associations between CF pulmonary function and mortality imply that pulmonary function is a valid surrogate end point, a position that has been supported by regulators [202]. FEV$_1$% predicted has been used to stratify both CF lung disease stage and disease aggressiveness [19,20], as well as CF care center performance [2,22]. An acute drop in FEV$_1$, 10–15% has been shown to be the most significant predictor of treatment for pulmonary exacerbation in patients 6 years and older [22], and treatment intensity has been correlated with an individual’s FEV$_1$% predicted [21,23,24].

Pulmonary function is most commonly employed as a CF clinical trial end point by quantitation of ‘sustained difference’ between randomized groups after fixed duration treatment with either ‘active’ drug or matched placebo under blinded conditions [25–37]. In simple terms, a study population is randomly divided into two groups, one is treated with active drug and the other placebo, and the mean pulmonary function of each group is compared after a fixed duration of treatment. This use of pulmonary function as an end point has been acceptable to regulatory agencies in support of registration of chronic CF respiratory therapies [202–209].

Many CF trials are conducted in growing subjects who may be experiencing progressive lung disease, creating a mathematical challenge for investigators studying treatment-associated differences in pulmonary function. Linear growth is accompanied by a net increase in FEV$_1$ in ‘healthy’ populations, while progressive lung disease can result in a relative loss of FEV$_1$ in subjects with CF. This problem can be addressed in clinical trials by normalization of FEV$_1$ for age, height and sex: a healthy child’s FEV$_1$% predicted should remain constant as their absolute FEV$_1$ increases with growth, while relative loss of FEV$_1$ during growth (which may still result in a modest increase in absolute FEV$_1$) would be detected as a decrease in FEV$_1$% predicted. In the past, normalization for age, height and sex has been most often achieved by converting FEV$_1$ to FEV$_1$% predicted using normative equations [38–41], and then studying differences in FEV$_1$% predicted among treatment groups [25–29]. Over time, choice of normative equations has changed as investigators have realized important distinctions in measures derived from different equations, particularly in children with relatively good lung function [42].

Regardless of the normative equation used, conversion of absolute FEV$_1$ to FEV$_1$% predicted can result in increased variance of measure, particularly when a subject’s age crosses ‘boundaries’ of normative equations during a study. Increased variance makes it more difficult to ascertain whether an observed difference in
means is a true difference or simply the result of random chance, leading to a requirement for relatively larger sample sizes and/or treatment effects to reach statistical significance. Growth potential creates a separate problem with FEV₁ as an end point in children and adolescents. An advantage to conversion of FEV₁ to FEV₁% predicted using normative equations is that adjustment for growth allows investigators to better characterize growth-independent changes in lung function.

Recently, CF investigators have noted that variance (and thus sample sizes) can be reduced by directly comparing differences in absolute FEV₁, while adjusting for height, age and sex during comparisons, rather than converting FEV₁ to FEV₁% predicted prior to analyses. This modification has been suggested to have the potential to reduce required sample sizes by as much as 15% for FEV₁ difference studies. Although this approach affords more precise differentiation between true treatment differences and random chance observations, the output unit of measure (adjusted liters) is not common to CF clinical practice, and may be more difficult to incorporate into risk/benefit analyses or to compare with previous CF clinical trial results.

A more fundamental challenge with sustained FEV₁ difference as a CF clinical trial end point is interpretation of its long-term clinical significance. Increasing an individual’s FEV₁% predicted may or may not change his or her future rate of pulmonary function deterioration (i.e., his or her disease aggressiveness), and it is the latter that has been proposed to be a more meaningful predictor of survival. To date, dornase alfa is the only chronic CF respiratory therapy that has been shown to both improve FEV₁ from baseline and alter disease course, as measured by rate of FEV₁ decline. By contrast, other chronic CF respiratory therapies (high-dose ibuprofen and inhaled corticosteroids) have been shown to slow the rate of pulmonary function decline over time without contributing a sustained FEV₁ increase.

An alternative end point to sustained FEV₁ difference (which captures treatment-related lung function improvement) is difference in mean rate of FEV₁ decline, which captures stabilization of lung function over time. Mean rate of FEV₁ decline has been used as a primary end point in several CF trials. Presumably, the extent to which a chronic therapy changes the rate at which FEV₁ is lost over time (as % predicted/year) is more clinically relevant than the simple difference in FEV₁ in the presence or absence of treatment. However, a correlation between decreasing rate of FEV₁ decline and increasing survival remains theoretical and might require a study decades in duration for true validation as a surrogate for increased survival.

FEV₁ decline is a more statistically demanding end point than sustained FEV₁ difference, because (unlike sustained difference) variances associated with mean decline rates are very sensitive to the duration over which they are measured (Figure 1). In other words, the longer a study population is observed, the more accurately their mean FEV₁ decline rate can be estimated. Because sample size requirements for rate of FEV₁ decline studies are inversely related to study duration, they can be relatively small when studies are conducted for years. In addition, limiting studies to subpopulations at higher risk for more rapid FEV₁ decline can further reduce sample size requirements. This inverse relationship between study duration and sample size requirements is not shared by FEV₁ difference end points, where the outcome measure may differ by time from baseline, but not in a predictable manner. Thus, duration of measure does not influence sample size calculations for studies employing FEV₁ difference end points. For example, the greatest (and thus the most statistically significant) differences in FEV₁% predicted observed between active and placebo groups in the 24-week inhaled tobramycin and oral (p.o.) macrolide studies were at 2 and 24 weeks, respectively. The rate of FEV₁ decline has been used successfully to demonstrate treatment efficacy of dornase alfa, high-dose ibuprofen and inhaled corticosteroids in retrospective analyses of CF patient registries.

**Frequency of pulmonary exacerbation**

CF lung disease progression is accompanied by increasingly frequent acute episodes of elevated respiratory signs and symptoms, termed pulmonary exacerbations, which require aggressive intervention. Pulmonary exacerbations are associated with reduced QoL and survival, and differences in relative risk of exacerbation or median time to next exacerbation have been employed as clinical end points for clinical trials of a variety of chronic CF therapies.

Unlike the pulmonary function end point, difference in risk of pulmonary exacerbation is a ‘true’ clinical end point that has supported the registration of several CF therapies. This is because an exacerbation is, in itself, a clinical event (affecting how a patient feels and functions). Although clinicians may interpret a reduction in risk of exacerbation as a reflective of an underlying change in lung disease or disease progression, it need not necessarily be so in order to remain a clinical benefit. However, several aspects of pulmonary exacerbation as an end point are challenging to investigators and regulators. For example, the underlying cause(s) of pulmonary exacerbations are not clear and they do not occur uniformly across the
exacerbations in these study designs, yet retrospective care centers of differing standards of care between clinicians investigators. First, variability is introduced in the form of exacerbation definition creates several problems for clinical criterion hospital admission and/or antibiotics as a diagnostic clinician’s decision to treat respiratory symptoms (with investigator’s limited analyses to subjects treated with iv. tobramycin investigators limited analyses to subjects treated with iv. antibiotics for exacerbations [25,27,28], while the p.o. macrolide study included subjects treated with p.o. antibiotics into exacerbation analyses [29].

There is also evidence that the overall ‘threshold’ for clinician intervention to treat exacerbation symptoms has shifted over recent decades, as evidenced by steadily improving mean pulmonary function [69] (which roughly inversely correlates with exacerbation incidence [56]) during a period where the incidence of iv. treatment for exacerbation has remained relatively constant in the US population [2]. For these reasons and others, use of exacerbation as a CF trial end point will remain challenging until a prospective definition of exacerbation becomes widely accepted.

QoL
Although recognized as a useful and important end point for chronic CF pulmonary therapies [13], measurements of patient-reported differences in QoL have only recently been used as a prospective efficacy end point in randomized CF clinical trials [31]. The tool used, the CF Questionnaire – Revised, measures CF health-related QoL on 12 generic and disease-specific scales [70], of which the respiratory symptom scale has been most emphasized in randomized clinical trials [71]. Although regulatory agencies have stressed the importance of patient-reported outcomes in measurement of treatment-associated benefits [202,206], there is a paucity of normative data with which to design, power or interpret results of prospective clinical trials using QoL end points in different CF populations, although this field is
Outcome measures for clinical trials assessing treatment of cystic fibrosis

Growing rapidly. Development of patient-reported tools in the context of measuring response to acute interventions (such as treatment of pulmonary exacerbation) have the potential to dramatically improve patient management [72].

Growth

There has yet to be a randomized prospective study of a pulmonary therapy in which difference in growth has been used as a primary efficacy end point, although there is good evidence that pulmonary health and growth are related in children with CF [73–76]. It is important to distinguish between weight-gain and linear growth as CF study end points. CF subjects can have low weight-for-age values and benefit from treatment-related weight-for-age increases at any age. However, the utility of linear growth as an end point is obviously limited to studies in infants, children and adolescents with linear growth potential. Differences in growth between treatment groups in randomized studies have been analyzed [26,47,77–79]. Growth has been used as a primary efficacy end point in a randomized controlled CF study of growth hormone in children, which demonstrated that significant improvements in stature associated with growth hormone treatment did not provide a pulmonary function benefit in children [80]. Growth may become a more useful clinical trial end point in the study of systemic therapies with the potential to produce increased CFTR [34,35,81–84] or CFTR-like activities, as multiorgan benefits of these therapies could include improved GI, endocrine and pulmonary function.

Near-global adoption of CF newborn screening [85,86] and broad adherence to nutritional management guidelines [87] has resulted in a much healthier cohort of CF children that are less likely to be nutritionally challenged [2]. Although improved nutritional health is clearly beneficial for CF children and families today, it reduces the ability of investigators to exploit growth difference (particularly in the form of BMI) as an efficacy measure in randomized trials, and it is unlikely that such a study will be (or could be) conducted in the future.

Shortcomings of current CF end points

Fundamental problems shared by the most commonly used end points described above are that they have their greatest utility in CF subpopulations with compromised pulmonary health [88], while our interests are centered more and more on chronic treatment to avoid compromised pulmonary health [89,90]. In general, these end points provide little or no value in evaluating the efficacy of pulmonary therapies in infants and younger children with ‘normal’ spirometric values, yet this is precisely the population that could benefit from intervention to prevent lung disease progression. Furthermore, improved pulmonary health [69] and survival [2] in the

---

Figure 2. Sample size requirements for 1:1 randomized exacerbation risk studies conducted in different age and FEV subgroups. Sample sizes per study arm required to detect a 40% reduction (hazard ratio: 0.6) in treatment for exacerbation with (A) inhaled antibiotics, oral fluoroquinolones or intravenous antibiotics and with (B) intravenous antibiotics. Values are provided for subgroups divided into age (abscissa axis) and FEV,% predicted (z-axis) subgroups. Estimates assume a 6-month blinded study, 1:1 randomization, 80% power and \(\alpha = 0.05\). Note that the ordinate scale in (B) is three-times the size of that of (A) [66].
CF population has decreased the size of the population that are best suited for study employing these endpoints [88,89], making it more difficult to test chronic pulmonary therapies. Pulmonary function endpoints require some pulmonary function loss in order to detect therapeutic benefit: either differences in sustained improvement in subjects who have already lost pulmonary function or in rates at which pulmonary function is lost. The risk of pulmonary exacerbation inversely correlates with FEV₁% predicted [56,66], such that studies employing exacerbation risk endpoints are not feasible in younger subjects with ‘normal’ FEV₁ (i.e., >90% of predicted; Figure 2) [66] and infants.

Unfortunately, FEV₁ loss does not signal the beginning of the CF lung disease process, but rather an intermediate stage of an advancing disease process. Lung function appears to be essentially normal in CF newborns [91], but thickened airway secretions and compromised MCC (an important component of lung defense after birth) immediately predispose infants to airway obstruction and infection that begins a life-long cycle of inflammation and destruction [12]. Perturbations of ‘normal’ lung physiology beyond depressed MCC detected in CF infants include gas trapping, bronchial wall thickening and dilation measured by high-resolution computed tomography (HRCT) [92–95], ventilation inhomogeneity measured by inert gas multiple breath washout [96–98] and reduced maximal expiratory flows measured by infant pulmonary function testing [99–101]. Simultaneously, early isolation of bacterial opportunists from the airway and detection of circulating antibodies to bacterial proteins [94,102] demonstrate a link between impaired MCC and increased risk of airway infection. Inflammatory biomarkers are elevated in respiratory secretions, particularly in children with evidence of local infection, confirming that a chronic inflammatory process and accelerated airway obstruction begin relatively early in life [94,103–108].

Epithelial ion transport, MCC, surface liquid dehydration, airway wall thickening, small airway obstruction, ventilation inhomogeneity, presence of opportunistic infection and inflammation are more than merely harbingers of more severe functional problems later in life, they also comprise the ‘therapeutic targets’ of both approved pulmonary therapies and those in development (Figure 3). Assays have been developed to assess the effect of interventions on most, if not all, of these CF-associated abnormalities in treated patients [109]. It is not unreasonable to hypothesize that interventions capable of mitigating these abnormalities in vivo early in the disease process (i.e., before loss of pulmonary function, increased risk of exacerbation or decreased QoL) may delay disease progression and subsequently increase survival. However, ‘upstream’ therapeutic targets shown in Figure 3 remain unvalidated biomarkers that may be useful to assess the potential for benefit of a new intervention early in the drug development process [109], despite observations that CF therapies proven to be clinically beneficial can influence them [110–115]. Although some CF clinicians may be convinced that affecting these biomarkers through intervention will inevitably lead to clinical benefit, regulators are tasked with quantifying intervention/clinical benefit relationships. Unless and until relationships between magnitude of treatment-related biomarker changes and the probability and nature of subsequent clinical benefit are fully described, it is not possible to objectively assess risks and benefits.

Figure 3. Cystic fibrosis lung disease therapeutic targets. Physiological ramifications of reduced CFTR activity in the cystic fibrosis lung are highlighted by large arrows. Therapeutic classes that have been and/or are being investigated for the chronic management of cystic fibrosis lung disease are shown in boxes, including CFTR gene therapy [127,128], small-molecule CFTR modulators [34,35,81–84], ion channel recruiters [36], hydrators [30,53], mucolytics [25], anti-infectives [27,31–33,52], Vaccines [129] and anti-inflammatories [26,47,50]. Despite different mechanisms of action, all share the goal of reducing lung disease damage caused by the interplay of obstruction, infection and inflammation.
Outcome measures for clinical trials assessing treatment of cystic fibrosis

Strengths & weaknesses of unvalidated ‘candidate’ end points
As noted previously, there are a number of measures that are potential candidates for end points in studies assessing the tangible benefits of pulmonary therapies. These measures are attractive to investigators because they are more directly associated with the underlying CF lung physiology and disease progression and/or have the potential to be assessed much earlier in the disease process. Although the underlying rationales supporting these end points are often sound and many have found utility in early proof-of-principle studies, practical considerations associated with their use as primary clinical efficacy end points in pivotal trials are rarely addressed. In many cases, determination of the magnitude of change of a given measure that can be considered to be ‘clinically meaningful’ has yet to be established. Without this information, justifications for and the feasibility of these outcomes as primary clinical trial end points remain unclear.

For example, lung imaging (e.g., HRCT) to characterize progressive structural damage has received enthusiastic academic support as a surrogate clinical end point \[116–120\]. Although HRCT is costly to investigators [121] and radiation exposure risks have been debated \[122,123\], the underlying rationale for this approach is nearly unassailable: end-stage CF lungs show massive disseminated structural damage \[124\], structural changes can be observed by HRCT in patients for which there is little or no evidence of spirometric change \[125\] and there is excellent evidence that bronchiectasis can begin in infancy \[126\]. It follows that an intervention that reduces or delays the occurrence of bronchiectasis or reverses bronchiectasis would modify CF lung disease progression and, by extension, measurement of change in bronchiectasis by imaging would be a valid study end point. However, with respect to a lung imaging end point, the devil lies in the details; the question is not a global one of whether development of bronchiectasis is clinically meaningful to an individual with CF (as it undoubtedly is), but one of quantitation. HRCT scoring algorithms are by their nature abstract, and the magnitude of change in score that is ‘clinically meaningful’ has yet to be unambiguously established. Longitudinal analyses linking early HRCT changes to later clinical outcomes (e.g., pulmonary function, QoL or survival) are required to establish clinical relevance of score changes, after which the feasibility of HRCT scoring as a study end point can be completely assessed. If natural history studies suggest that clinically meaningful changes in HRCT scores occur only over extended time periods, infrequently in a given population, or both, then sample size considerations could render the end point impractical, regardless of clinical relevance. By extreme analogy, mortality is a clinically relevant but impractical CF clinical trial end point due to its relative infrequency in the population.

The recent use of more ‘direct’ measures of CFTR protein function (e.g., change in sweat chloride and/or nasal potential difference \[34,35,81–84\]) as biomarkers of efficacy for small-molecules targeting the primary CF defect have been encouraging, in that they suggest that these types of measures may soon be validated as surrogates for assessment of clinical benefit. Although their utility is limited to the study of systemic therapies where CFTR function is also affected outside of the lower airway (e.g., in the sweat gland or the nares), these biomarkers appear to be particularly robust because of their rapid change in the presence of treatment and our appreciation of the role of CFTR dysfunction in the etiology of CF lung disease. However, as with other candidate biomarkers, validation of change in sweat chloride or nasal potential difference as independent surrogates of clinical efficacy will require quantitation of relationships between biomarker values or changes in their values, and the probability and magnitude of subsequent clinical benefits. While there is reason for optimism that this may be accomplished in the future, it has yet to occur.

Summary
Demonstration of clinical efficacy for chronic CF respiratory therapies has relied most heavily on two measures relevant to patients with some degree of airway damage: pulmonary function (as measured by FEV1) and pulmonary exacerbation. Patient-reported QoL has only recently been used as a CF end point, but it too may be best suited for patients with airway impairment. Growth has not been particularly useful as a trial end point, and the advent of CF newborn screening and improved nutritional management suggest that it may never be.

Ironically, CF therapies generally target aspects of CF pathophysiology that are present long before lung function, as measured by spirometry, becomes impaired. As clinicians become committed to arresting airway disease progression before substantial damage has occurred, current clinical trial end points become less useful, and there is great interest in validating underlying physiologic biomarkers as clinical trial end points. A number of upstream biomarkers have been described as candidates for surrogate clinical end points and the underlying rationales for
## Executive summary

### Cystic fibrosis
- A chronic disease dominated by inexorable progressive lung damage.
- Chronic respiratory therapies target underlying pathophysiologic changes in the cystic fibrosis (CF) airway.
- Demonstration of treatment efficacy is most often achieved using one or more of three different measures:
  - Improvement/stabilization of pulmonary function;
  - Frequency of pulmonary exacerbation;
  - Quality of life (QoL).
- Growth has been identified as a valid clinical end point.
- A variety of biomarkers have been used as end points in proof-of-concept studies:
  - Biomarkers have yet to be accepted as clinical surrogates by regulators.

### Improvement/stabilization of pulmonary function
- Sustained improvement in pulmonary function is measured by cross-sectional comparison of change from baseline for treatment arms:
  - Commonly employed in drug registration studies;
  - Difficult to interpret long-term clinical significance.
- Slowing in rate of pulmonary function decline compares mean rates of lung function decline between treatments:
  - More challenging end point to employ;
  - More compelling clinical significance; suggestive of disease modification.

### Frequency of pulmonary exacerbation
- True clinical end point, important to registration of several chronic CF therapies.
- More difficult to power than sustained improvement in pulmonary function.

### QoL
- True clinical end point, only recently emphasized.
- Normative data required for the design, powering and interpretation of trials using QoL end points in development.

### Growth
- Yet to be used as a primary end point for a respiratory therapy.
- Greater potential for use with systemic therapies that modulate mutant CFTR activity.
- Newborn screening and greater attention to nutritional support have reduced the potential signal for growth as an end point in infants.

### Shortcomings of current CF end points
- Applicability to that fraction of CF patients with more advanced airway disease.
- Ambiguity as to what magnitude of change/difference constitute clinically meaningful treatment effects.
- For exacerbation, lack of a consensus definition of exacerbation.

### Strengths & weaknesses of unvalidated ‘candidate’ end points
- Biomarker end points are attractive measures:
  - More direct association with underlying CF lung physiology and disease progression;
  - Can be assessed much earlier and more rapidly in the disease process.
- Validation of biomarkers as clinical surrogates remains incomplete:
  - Incomplete description of relationships between change in marker values and subsequent changes in clinical end points;
  - Lack of descriptions of ‘clinically meaningful’ changes in marker values.
- Insufficient information to assess the practicality of biomarkers as surrogate clinical surrogates:
  - Frequency and rate of clinically meaningful changes in the biomarkers within populations are required to assess sample size and study duration requirements.

### Summary
- Currently accepted clinical end points for CF pulmonary therapies are best-suited for subjects with either very aggressive or established lung disease:
  - As the health of the CF population improves and clinicians hope to intervene to prevent disease progression, these end points become less useful in identifying disease-modifying therapies.
- Continued advance in the management of CF will require development and use of clinical trial end points applicable earlier in life and in the airway disease process.
- Validation of upstream biomarkers as clinical surrogates will require characterization of the relationships between changes in biomarkers and subsequent clinical outcomes.
their use are often solid. Indeed, many of these biomarkers have already been used in early-stage proof-of-concept studies for pulmonary therapies. However, validation of these biomarkers as clinical surrogates, which could greatly improve our ability to treat the earliest stages of CF lung disease progression, will require careful documentation of relationships between treatment-associated changes in biomarkers and subsequent changes in clinically valid end points, such as pulmonary function, QoL or survival.

Future perspective
Improving health of the CF population and emphasis on disease prevention will put greater pressure on investigators and regulators to develop and validate early biomarkers of pulmonary disease progression as clinical trial end points for the registration of chronic CF respiratory therapies. Validation of ‘investigational’ biomarkers of lung health, such as MCC, ventilation homogeneity, inflammatory status and imaging through correlation of biomarker changes early in CF lung disease with currently accepted CF clinical end points (pulmonary exacerbation, pulmonary function, QoL and survival) will require a concerted effort among CF caregivers to contribute accurate and complete data to patient registries and support rigorous longitudinal studies to characterize relationships between upstream biomarkers and subsequent clinical status.

Financial & competing interests disclosure
This work was funded by the gracious support of the NIH National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Grant P30 DK027651. The content is solely the responsibility of the authors and does not represent the views of the NIDDK or the NIH. DR VanDevanter has served as a CF advisor or consultant in the past year for the Cystic Fibrosis Foundation, Inspire Pharmaceuticals, Genentech, Gilead Sciences, KaloBios Pharmaceuticals, Nanobio Corporation, Pulmatrix, Rempex Pharmaceuticals, and Vertex Pharmaceuticals. MW Konstan has served as a CF advisor or consultant in the past year for Abbott Laboratories, Axxon Pharma, Aradigm Corporation, Boehringer-Ingelheim, Digestive Care, Genentech, Gilead Sciences, GlaxoSmithKline, Inspire Pharmaceuticals, KaloBios Pharmaceuticals, Eli Lilly, Novartis, PTC Therapeutics, Inomed, Vertex Pharmaceuticals, and CSL-Behring. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

References
Papers of special note have been highlighted as:
- of interest


This article employs retrospective analysis of a large cystic fibrosis (CF) registry to describe nuances of rate of FEV1 decline as a CF clinical trial end point.


Outcome measures for clinical trials assessing treatment of cystic fibrosis

Review: Clinical Trial Methodology


This article employs retrospective analysis of a large CF registry to describe nuances of pulmonary exacerbation as a clinical trial end point.


Describes a widely-used and validated health-related quality of life measurement tool developed specifically for CF.


### A comprehensive summary of CF clinical end points.


### Websites

201 Cystic fibrosis mutation database. www.genet.sickkids.on.ca/StatisticsPage.html


