



# Systematic review of content and quality of idiopathic pulmonary fibrosis review articles

Kathryn M. Milne <sup>1</sup>, Chrystal Chan <sup>1</sup>, Jolene H. Fisher <sup>2</sup>, Kaïssa de Boer <sup>3</sup> and Christopher J. Ryerson <sup>1,4</sup>

**Affiliations:** <sup>1</sup>Dept of Medicine, University of British Columbia, Vancouver, BC, Canada. <sup>2</sup>Dept of Medicine, University of Toronto, Toronto, ON, Canada. <sup>3</sup>Dept of Medicine, University of Ottawa, Ottawa, ON, Canada. <sup>4</sup>Centre for Heart Lung Innovation, University of British Columbia and St. Paul's Hospital, Vancouver, BC, Canada.

**Correspondence:** Christopher J. Ryerson, Dept of Medicine, University of British Columbia – St. Paul's Hospital, 1081 Burrard St, Ward 8B, Vancouver, V6Z 1Y6, Canada. E-mail: [chris.ryerson@hli.ubc.ca](mailto:chris.ryerson@hli.ubc.ca)

**ABSTRACT** Narrative reviews are frequently accessed; however, the content and quality of review articles on idiopathic pulmonary fibrosis (IPF) have not been assessed.

A systematic review assessed content and quality of narrative review articles that addressed the diagnosis or management of IPF and were published from 2001 to 2015. Article recommendations were assessed relative to contemporary IPF guidelines. Quality was assessed using the DISCERN instrument.

Articles were predominantly written by physicians and published in respiratory journals. Conflicts of interest and sources of funding were reported in 52% and 24% of reviews, respectively. European authors were more likely to recommend bronchoscopy (adjusted  $p=0.02$ ) and were more likely to recommend pirfenidone or nintedanib prior to publication of definitive clinical trials (adjusted  $p=0.04$ ). A total of 39% of management-focused articles suggested therapies that were never recommended in guidelines. Predictors of higher article quality were citation of the contemporary IPF guideline ( $p=0.01$ ) and more recent publication ( $p=0.001$ ).

Quality of reviews increased over time; however, review articles frequently made discordant recommendations compared to IPF guidelines. These findings indicate the need for authors, peer reviewers, editors and readers to critically appraise the content and quality of narrative reviews on IPF, and the need for frequent guideline updates to reflect new evidence.



@ERSpublications

**Recommendations for the diagnosis and management of IPF differ between North America and Europe and review articles frequently make recommendations for management of IPF that are discordant from international guidelines despite increased quality over time**

<http://ow.ly/KeLh30mNLUt>

**Cite this article as:** Milne KM, Chan C, Fisher JH, *et al.* Systematic review of content and quality of idiopathic pulmonary fibrosis review articles. *ERJ Open Res* 2018; 4: 00156-2018 [<https://doi.org/10.1183/23120541.00156-2018>].



## Introduction

Idiopathic pulmonary fibrosis (IPF) is a chronic progressive fibrotic lung disease that is characterised by progressive dyspnoea, reduced quality of life and early mortality. Diagnostic criteria for IPF have evolved since publication of the first major international guideline in 2000 [1, 2]. Management of IPF has also changed substantially over the past decade [2, 3], including evidence that previously employed therapies increase mortality [4], and with the approval of new antifibrotic medications that slow disease progression [5–7]. These rapid advances highlight the importance of ensuring that healthcare providers are accessing accurate and high-quality literature on the diagnosis and management of IPF.

Narrative review articles are highly accessed educational resources that are often viewed more frequently than original research publications [8]. Despite this high visibility, the content of narrative review articles is infrequently assessed, and there are no previous studies that have addressed the content or quality of review articles on the diagnosis and management of IPF. We therefore conducted a novel systematic review to determine the content and quality of narrative review articles that address the diagnosis or management of IPF. This study was also performed with the intent of providing detailed and reproducible methods for this process to be undertaken in other chronic diseases.

## Methods

### *Data sources and search strategy*

A systematic search of the Medline and EMBASE databases was performed to identify all narrative review articles published in English between 2001 and 2015 that addressed the diagnosis or management of IPF. The date range was selected to encompass review articles published following the two international IPF clinical practice guidelines in 2000 and 2011 [1, 2], with the end date of 2015 selected to precede publication of the 2015 guideline update on the management of IPF [3]. Two authors (K.M. Milne and C. Chan) independently reviewed and extracted data using pre-defined criteria. Our systematic review was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist where applicable [9]. Our systematic review was not registered with PROSPERO (the international prospective register of systematic reviews) as we did not assess any health-related outcomes. Details of the search strategy, developed in consultation with a librarian, are available in supplementary table S1.

### *Review selection and data extraction*

Narrative review articles that addressed the diagnosis or management of IPF in human adults were included. Two authors independently reviewed titles and abstracts for eligibility criteria. Publications selected for full-text review were then assessed by two authors to ensure eligibility. All narrative review articles published in English, between 2001 and 2015, addressing the diagnosis or management of IPF in adults were eligible for inclusion. Systematic reviews were excluded given differences in methodology between systematic reviews and narrative reviews, limiting comparison between their recommendations. Data from included reviews were abstracted using pre-defined data extraction forms that included information related to article objective, journal-related variables, author-related variables, concordance with guideline recommendations, quality assessment, and citation of guidelines and landmark trials (supplementary table S2). Differences were resolved by iteration and consensus with a third reviewer resolving any remaining disagreements.

### *Content and quality assessment*

Diagnosis and management recommendations made within each review article were compared to recommendations made in the contemporary guideline (*i.e.* review articles published between 2001 and 2011 were compared to the 2000 guideline [1], and review articles published between 2012 and 2015 were compared to the 2011 guideline) [2]. Recommendations made in each guideline are summarised in supplementary table S3. Affirmative recommendations made within a review article were defined as those that were presented as routine diagnostic tests in the evaluation of IPF or routine therapies that should generally be used in most IPF patients. Articles were assessed for concordant and discordant recommendations related to management of IPF compared to contemporary international IPF guidelines, recognising that guidelines can become outdated and only more recently have adhered to more rigorous methodology and grading standards. The total number of concordant and discordant recommendations in each review article was calculated. The timing of review article recommendations was also compared to the 2014 publication of the ASCEND and INPULSIS trials for pirfenidone and nintedanib [5, 7], and to the 2011 publication of the CAPACITY trials for pirfenidone only [6].

The quality of written information on treatment choices was assessed using the validated DISCERN instrument [10]. The DISCERN instrument is used to critically appraise and assess the quality of written health information for any disease, and is thus a useful tool for identifying unbiased, current and

comprehensive health information [10]. The DISCERN instrument is not intended to assess content itself. The DISCERN instrument consists of 16 questions rated from 1 (low quality) to 5 (high quality). The first eight DISCERN items assess publication reliability, items 9–15 assess management recommendation quality and item 16 is a global assessment. Two authors rated each item, with DISCERN scores within 1 point considered to be in agreement. Disagreement of >1 point prompted reassessment of the publication by the primary reviewing author, with further discussion as necessary to achieve consensus. Reviewers were not blinded to the journal of publication, but the DISCERN instrument was applied prior to extracting journal- and author-specific variables (e.g. impact factor, number of publications) in an attempt to minimise bias. The DISCERN instrument has previously been used in a similar manner to evaluate the quality of Internet-based written information on IPF [11].

### Statistical analysis

Data are reported as number with percentage, mean±SD or median with interquartile range. Unadjusted correlations between variables were assessed using a Chi-squared test, Fisher's exact test, t-test, Wilcoxon rank-sum test or Spearman correlation, as appropriate. Logistic regression was used to determine predictors of individual recommendations for the diagnosis and management of IPF. Poisson regression was used to determine predictors for the total number of concordant or discordant management recommendations made in each review article that focused on IPF management. Stepwise linear regression was used to identify independent predictors of article quality (DISCERN score), considering variables with an unadjusted  $p < 0.10$  and retaining variables in the final model with an adjusted  $p < 0.05$ . A two-sided  $p$ -value threshold of 0.05 was used to define statistical significance. Analyses were performed using STATA 11.2 (StataCorp, College Station, TX, USA).

## Results

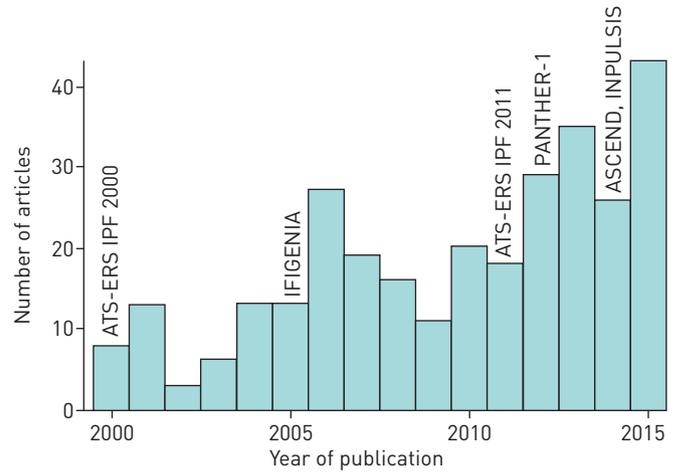
### Summary of included articles

Of 1148 unique citations identified in the search, 292 full-text articles were eligible and included in the analysis (table 1 and supplementary figure S1). Articles were predominantly written by physicians (92%), addressed management of IPF (76%), and were published in respiratory journals (62%). An impact factor was available for 185 articles (median (interquartile range) impact factor of 2.8 (2.0–4.0)). Conflicts of interest and sources of funding were reported in 52% and 24% of reviews, respectively. The number of articles published per year increased over time, particularly following publication of major international

TABLE 1 Description of included articles

<b>Country of corresponding author</b>	
North America	150 (51%)
Europe	105 (36%)
Other	37 (13%)
<b>MD versus other degree for corresponding author</b>	268 (92%)
<b>Conflict of interest</b>	
Present	92 (32%)
Absent	61 (21%)
Not reported	139 (48%)
<b>Funding source</b>	
Industry	21 (7%)
Non-industry	41 (15%)
None	5 (2%)
Not reported	225 (76%)
<b>Number of IPF publications by corresponding author</b>	2 [0–8]
<b>Journal type</b>	
Respiratory	181 (62%)
General internal medicine	38 (13%)
Pharmaceutical science	35 (12%)
Other	38 (13%)
<b>Journal impact factor</b>	2.8 [2.0–4.0]
<b>Primary objective</b>	
Diagnosis of IPF	161 (55%)
Management of IPF	221 (76%)
Diagnosis and management of IPF	90 (31%)

Data are presented as n (%) or median (interquartile range). MD: medical doctor; IPF: idiopathic pulmonary fibrosis.



**FIGURE 1** Number of idiopathic pulmonary fibrosis (IPF) review articles published per year from 2000 to 2015. See supplementary table S2 for complete reference of notable IPF-related publications. ATS-ERS IPF: American Thoracic Society/European Respiratory Society IPF guidelines; IFIGENIA: trial of *N*-acetylcysteine [12]; PANTHER-1: trial of prednisone, azathioprine and *N*-acetylcysteine [4]; ASCEND: trial of pirfenidone [5]; INPULSIS: trial of nintedanib [7].

guidelines and landmark clinical trials (figure 1). The 2000 IPF guideline was cited by 66% of review articles published between 2001 and 2011, and the 2011 guideline was cited by 85% of articles published between 2012 and 2015. The 2000 guideline was cited by 29% of management articles published between 2012 and 2015, including 5% that did not cite the 2011 guidelines.

**Diagnostic recommendations**

Diagnostic recommendations were provided in 161 reviews (table 2). Multidisciplinary discussion for the diagnosis of IPF was recommended in 60 articles (37% of all articles making any diagnostic recommendation), with no significant differences identified between reviews that did and did not provide this recommendation. Consideration of a surgical lung biopsy was recommended for the diagnosis of IPF in 103 articles (64%). On unadjusted analysis, recommendation for surgical lung biopsy was associated with North American authorship, earlier year of publication, fewer IPF-related publications by the corresponding author, and the inclusion of a conflict of interest statement. An earlier year of publication was the only independent predictor of a recommendation for surgical lung biopsy ( $p < 0.001$ ). Bronchoalveolar lavage and/or transbronchial biopsy were recommended in 42 articles (26%). Articles with an earlier year of publication ( $p = 0.04$ ) and European authorship ( $p = 0.02$ ) were more likely to recommend bronchoscopy for the diagnosis of IPF on adjusted analysis.

**TABLE 2** Unadjusted predictors of recommendation for multidisciplinary discussion, surgical lung biopsy and bronchoscopy

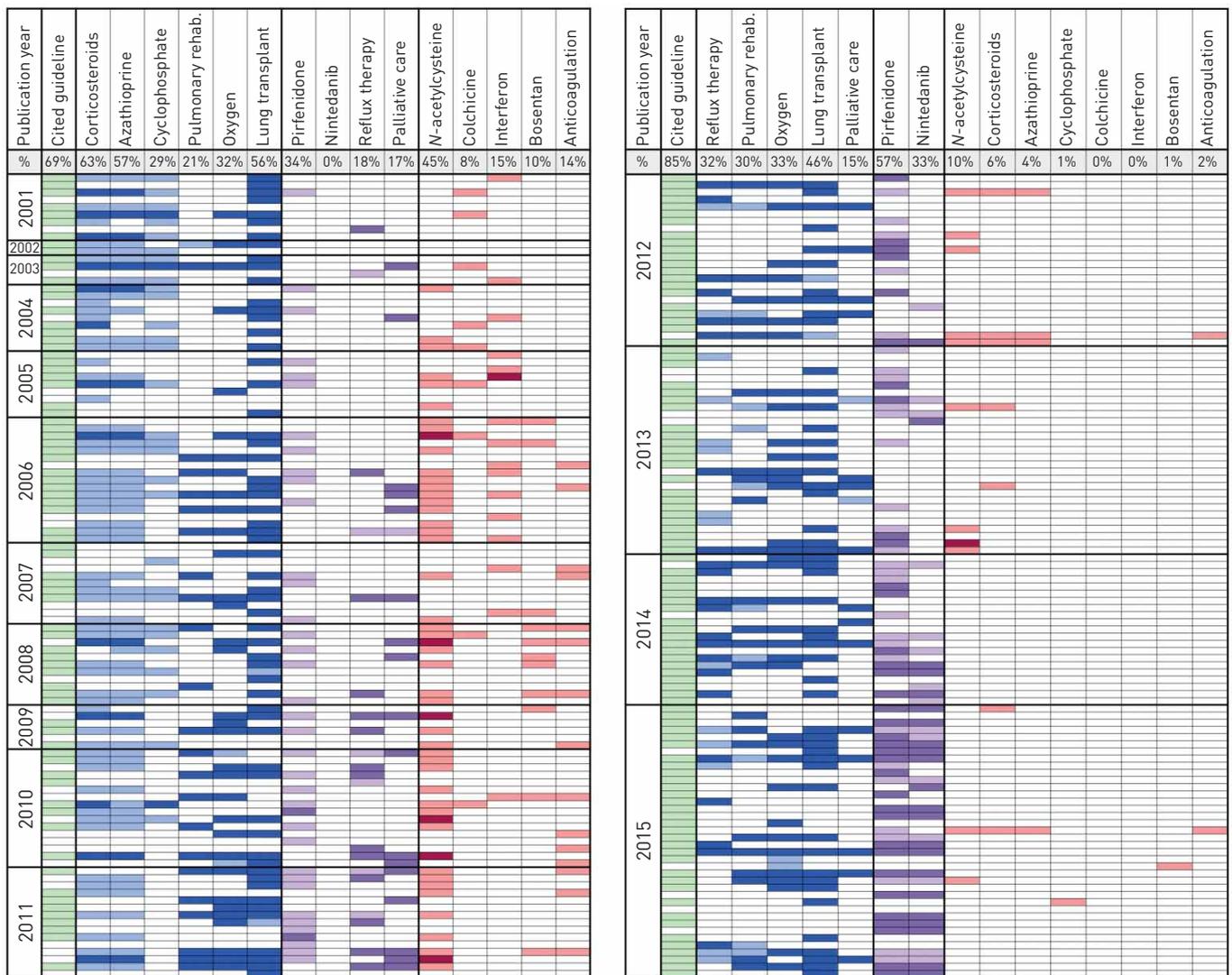
Predictor variables	Multidisciplinary discussion		Surgical lung biopsy		Bronchoscopy	
	OR (95% CI)	p-value	OR (95% CI)	p-value	OR (95% CI)	p-value
NA versus European corresponding author <sup>#</sup>	1.09 [0.60–1.97]	0.79	1.84 (1.09–3.11)	0.02	0.51 [0.26–1.02]	0.06
Year of publication <sup>#,¶</sup>	1.01 [0.95–1.09]	0.67	0.86 [0.81–0.92]	<0.001	0.92 [0.86–0.99]	0.04
MD versus other degree for corresponding author	1.37 [0.45–4.18]	0.58	2.28 [0.82–6.34]	0.11	0.63 [0.22–1.80]	0.39
Conflict of interest reported	0.82 [0.47–1.42]	0.47	0.46 [0.28–0.75]	0.002	0.69 [0.36–1.30]	0.25
Funding sources reported	1.22 [0.65–2.31]	0.53	0.87 [0.49–1.52]	0.62	1.39 [0.68–2.82]	0.37
Number of IPF publications by corresponding author	1.01 [0.98–1.04]	0.59	0.96 [0.93–0.99]	0.02	0.97 [0.92–1.02]	0.18
Respiratory journal	0.94 [0.53–1.67]	0.85	1.03 [0.63–1.68]	0.91	1.13 [0.58–2.20]	0.71
Journal impact factor, available	0.88 [0.50–1.55]	0.65	1.21 [0.73–1.98]	0.46	0.55 [0.29–1.04]	0.07
Journal impact factor, raw value	1.00 [0.92–1.08]	0.99	0.96 [0.88–1.04]	0.29	0.93 [0.79–1.11]	0.43
Cited contemporary guideline	1.65 [0.83–3.30]	0.15	0.86 [0.50–1.47]	0.58	1.08 [0.52–2.26]	0.84
DISCERN score <sup>+</sup>	1.00 [0.95–1.05]	0.94	0.98 [0.95–1.02]	0.34	0.97 [0.92–1.03]	0.30

NA: North American; MD: medical doctor; IPF: idiopathic pulmonary fibrosis. <sup>#</sup>: independent predictors of recommendation for bronchoscopy; <sup>¶</sup>: independent predictor of recommendation for surgical lung biopsy; <sup>+</sup>: per 10-unit change in DISCERN score.

**Management recommendations**

Management recommendations were provided in 221 reviews (figure 2). The number and strength of management recommendations made in narrative reviews between 2001 and 2015 are shown in figure 2, illustrating that more recent publications are less likely to suggest therapies that were never recommended in guidelines and that pirfenidone and nintedanib were frequently suggested in narrative reviews prior to their recommendation in guidelines. For articles published between 2001 and 2011, the total number of concordant management recommendations in each review was associated with higher DISCERN score and citation of the 2000 guideline (table 3). On unadjusted analysis of review articles published between 2012 and 2015, the number of concordant recommendations for management in each review was associated with higher DISCERN score and lower number of IPF-related publications by the corresponding author (table 3). Higher DISCERN score was the only independent predictor of the total number of concordant recommendations for both time periods ( $p=0.007$  and  $0.02$ , respectively).

Between 2001 and 2011, 76% of review articles recommended a treatment not supported in the 2000 IPF guidelines (*i.e.* discordant recommendations). From 2012 to 2015, 79% of articles made treatment recommendations discordant from the 2011 IPF guidelines. A total of 39% of management articles suggested pharmacotherapies that were not recommended in any IPF guideline, including the 2000, 2011



**FIGURE 2** Concordance of review article recommendations with contemporary guidelines. Each row represents one review article. Darker colours represent strong recommendations for the specified therapy and lighter colours represent weak recommendations. Cells shown in blue represent articles that provided a concordant recommendation for a treatment that was also recommended in the contemporary guideline. Cells shown in purple represent articles that provided a recommendation for a treatment that was not recommended in the contemporary guideline, but that was recommended in a subsequent guideline. Cells shown in red represent articles that provided a discordant recommendation for a treatment that was not recommended in the contemporary guideline or in a subsequent guideline. Rehab: rehabilitation.

TABLE 3 Unadjusted predictors of a concordant review article recommendation compared to the contemporary guideline

Predictor variables	2001–2011		2012–2015	
	Coefficient (95% CI)	p-value	Coefficient (95% CI)	p-value
NA versus European corresponding author	−0.19 (−0.51–0.14)	0.26	0.26 (−0.12–0.63)	0.18
Year of publication	0.001 (−0.05–0.05)	0.97	0.01 (−0.16–0.18)	0.91
MD versus other degree for corresponding author	0.31 (−0.46–1.09)	0.43	−0.17 (−0.61–0.26)	0.43
Conflict of interest reported	−0.11 (−0.42–0.20)	0.49	0.03 (−0.37–0.43)	0.88
Funding sources reported	−0.28 (−0.71–0.15)	0.20	−0.23 (−0.66–0.21)	0.31
Number of IPF publications by corresponding author	−0.01 (−0.05–0.03)	0.60	−0.02 (−0.04–0)	0.05
Respiratory journal	0.12 (−0.19–0.44)	0.44	−0.16 (−0.56–0.23)	0.42
Journal impact factor, available	−0.07 (−0.40–0.26)	0.69	0 (−0.38–0.38)	1.00
Journal impact factor, raw value	−0.01 (−0.06–0.04)	0.68	−0.03 (−0.11–0.04)	0.42
Cited 2000 guideline	0.43 (0.09–0.76)	0.01	0.38 (−0.01–0.77)	0.06
Cited 2011 guideline			0.05 (−0.55–0.66)	0.86
DISCERN score <sup>#,¶</sup>	0.26 (0.07–0.45)	0.007	0.37 (0.07–0.68)	0.02

All coefficients are based on a Poisson analysis. NA: North American; MD: medical doctor; IPF: idiopathic pulmonary fibrosis. #: independent predictor of concordant recommendation for 2001–2011 and 2012–2015; ¶: per 10-unit change in DISCERN score.

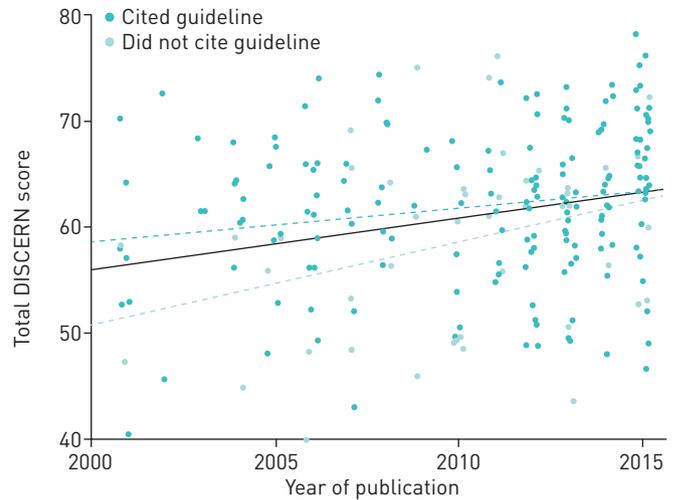
and 2015 documents [1–3]. On unadjusted analysis of publications between 2001 and 2011, the only predictor of the number of discordant management recommendations was more recent year of publication (table 4). On unadjusted analysis of publications between 2012 and 2015, a higher number of discordant treatment recommendations was associated with more recent year of publication, higher impact factor, publication in a respiratory journal, and citation of the earlier 2000 IPF guideline (table 4). On adjusted analysis, more recent publication year was the only independent predictor of a higher number of discordant recommendations for both time periods ( $p \leq 0.001$ ).

Recommendations that were discordant with the contemporary guideline increased over time, but were primarily recommendations for pirfenidone or nintedanib that pre-dated their support in the 2015 guideline ( $p < 0.001$  for trend over time). Recommendation for pirfenidone or nintedanib prior to the publication of definitive clinical trials in 2014 was more common with European authorship and a higher number of previous IPF-related publications by the corresponding author ( $8.0 \pm 7.8$  versus  $4.2 \pm 6.2$ ;  $p < 0.001$ ). Both variables remained independent predictors on adjusted analysis ( $p = 0.04$  and  $p < 0.001$ , respectively). Findings were unchanged when considering the 2011 CAPACITY trials as the definitive publication supporting use of pirfenidone.

TABLE 4 Unadjusted predictors of a discordant review article recommendation compared to the contemporary guideline

Predictor variables	2001–2011		2012–2015	
	Coefficient (95% CI)	p-value	Coefficient (95% CI)	p-value
NA versus European corresponding author	−0.15 (−0.50–0.20)	0.41	−0.02 (−0.30–0.26)	0.88
Year of publication <sup>#</sup>	0.13 (0.07–0.18)	<0.0005	0.20 (0.08–0.33)	0.001
MD versus other degree for corresponding author	−0.03 (−0.60–0.54)	0.92	0.12 (−0.25–0.50)	0.52
Conflict of interest reported	0.17 (−0.20–0.54)	0.38	0.02 (−0.32–0.37)	0.89
Funding sources reported	−0.12 (−0.57–0.33)	0.60	−0.03 (−0.33–0.27)	0.84
Number of IPF publications by corresponding author	0.02 (−0.01–0.05)	0.25	0.01 (0–0.18)	0.11
Respiratory journal	−0.17 (−0.55–0.21)	0.38	−0.32 (−0.58–−0.06)	0.02
Journal impact factor, available	−0.34 (−0.71–0.04)	0.08	−0.07 (−0.35–0.20)	0.60
Journal impact factor, raw value	0.01 (−0.03–0.05)	0.53	0.04 (0.01–0.08)	0.02
Cited 2000 guideline	0.26 (−0.15–0.67)	0.21	0.46 (0.18–0.74)	0.001
Cited 2011 guideline			−0.21 (−0.52–0.08)	0.15
DISCERN score <sup>¶</sup>	0.12 (−0.12–0.37)	0.33	0.11 (−0.07–0.29)	0.23

All coefficients are based on a Poisson analysis. NA: North American; MD: medical doctor; IPF: idiopathic pulmonary fibrosis. #: independent predictor of discordant recommendation for 2001–2011 and 2012–2015; ¶: per 10-unit change in DISCERN score.



**FIGURE 3** Total DISCERN score versus publication year, stratified by whether the review article cited or did not cite the contemporary international idiopathic pulmonary fibrosis guideline.

### Article quality

On unadjusted analysis, predictors of higher article quality (DISCERN score) for reviews addressing the diagnosis of IPF included impact factor availability, number of IPF-related publications by the corresponding author, and inclusion of a funding source statement (supplementary table S4). Impact factor availability was the only independent predictor of higher DISCERN score for reviews addressing the diagnosis of IPF ( $p=0.003$ ). On unadjusted analysis, predictors of higher article quality (DISCERN score) for reviews addressing the management of IPF included more recent year of publication (figure 3), higher number of IPF-related publications by the corresponding author, inclusion of a conflict of interest statement and citation of the contemporary guideline (supplementary table S4). Independent predictors of higher quality for review articles addressing management of IPF were citation of the contemporary guideline ( $p=0.01$ ) and more recent year of publication ( $p=0.001$ ); however, these variables explained only a small amount of variability in the DISCERN score ( $R^2=0.08$ ).

### Discussion

We conducted a novel systematic review to evaluate the content and quality of narrative review articles that address the diagnosis or management of IPF. This study shows that diagnostic and therapeutic recommendations made in narrative reviews have evolved over time (as expected), with frequent discordance compared to international clinical practice guidelines. We found that surgical lung biopsy and bronchoscopy were less frequently recommended for the diagnosis of IPF in recent years, while recommendations for pirfenidone and nintedanib in narrative reviews frequently preceded their endorsement in clinical practice guidelines and also preceded publication of definitive clinical trials. Furthermore, we show differences in diagnostic and therapeutic recommendations made by North American compared to European authors and have identified the challenges in assessing the quality of narrative reviews.

We observed a significant amount of discordance between recommendations made in guidelines and those made in narrative reviews, despite the use of more rigorous standards in more recent guidelines. One issue that has been difficult to adequately address in guidelines is a rapidly changing evidence base that quickly results in a guideline becoming outdated. This highlights the utility of guidelines with frequent updates such as those from the Global Initiative for Asthma [13]; however, a rapidly changing evidence base fails to account for all of the discordance observed in our study. Importantly, there is frequent subjectivity to the recommendations made in guidelines, with final recommendations typically based on majority vote. Such votes are often not unanimous, reflecting differences in opinion among guideline committee members and leading to some uncertainty as to whether guidelines are indeed the gold standard that they are often portrayed as being. Our findings highlight this fact, showing significant discordance between recommendations made by authors in narrative review articles compared to recommendations made by a guideline committee that often includes these same experts. For this reason, we refrained from labelling guidelines as “truth”, and instead our analyses are based on the identification of simple discordance without designating either guidelines or narrative reviews as correct or incorrect.

The roles of both bronchoscopy and surgical lung biopsy in the diagnosis of IPF have evolved since the 2000 IPF guideline. The less frequent recommendation for surgical lung biopsy over time in narrative reviews is likely to be due to greater appreciation of surgical risks and improvements in the imaging-based

diagnosis of IPF. Bronchoscopy remains controversial, with some authors recommending bronchoscopic evaluation in patients with suspected IPF and some recommending against this. Recommendations distinguishing between traditional transbronchial biopsy and cryobiopsy were not extracted in our systematic review. We found that European authors more frequently suggested bronchoalveolar lavage in assessing patients for IPF compared to North American authors. This geographic difference has been highlighted in previous perspectives and pro-con debates [14, 15]. Our study provides the first objective data that these differences are reflected in the published literature, which then probably filters down to influence the practice patterns of the clinicians who read these articles. These geographic differences also highlight some of the challenges with achieving consensus within diverse clinical practice guideline panels.

We similarly found geographic differences in the frequency of recommendation for pirfenidone or nintedanib prior to the publication of major clinical trials that supported these therapies. The conflicting but generally favourable results of the two CAPACITY trials of pirfenidone resulted in approval of the drug in Europe and many other regions, but not in the USA [6]. This earlier access to pirfenidone in Europe coincided with the more frequent recommendation for this therapy from European authors prior to publication of the ASCEND trial [5]. Interestingly, there were similar geographic differences in the recommendation for nintedanib prior to the INPULSIS trials being published [7]. Multiple countries of origin for corresponding authors of narrative reviews included in our study have published national or regional IPF guidelines. We did not assess such national guidelines given the different languages of publication and our inclusion of narrative reviews published only in English. It is unknown whether local factors (e.g. unique national guidelines, decisions of regulatory agencies, local availability of medication) could have influenced the views of local authors, whether the views of local key opinion leaders influenced decisions made by these agencies, or whether these occurred in concert due to other factors.

Our novel approach to assessment of the literature is particularly important in diseases like IPF that have a rapidly evolving evidence base that quickly leads to outdated recommendations being made in highly visible review articles. Authors should incorporate criteria provided in assessment tools such as DISCERN to increase the quality of a narrative review through several simple mechanisms, such as disclosing conflicts of interest and funding sources, which we found were frequently omitted. Assessment of a review article's content is often more challenging. Of specific concern, we found that 39% of reviews recommended a therapy that has never been supported in the international guidelines, a particularly worrisome finding given the challenges readers face in evaluating the quality of these narrative reviews. This emphasises the need for authors of narrative reviews to carefully consider whether an evolving evidence base supports any departure from recommendations made in previous clinical practice guidelines, again recognising that these guidelines are not unimpeachable gold standards.

In summary, this systematic review identifies the limitations of narrative review articles that address the diagnosis and management of IPF. We highlight the challenges in assessing article quality by showing that there is no reliable way by which readers can determine article quality. We furthermore demonstrate significant regional differences comparing recommendations made by North American and European authors in the diagnosis and management of IPF. These findings indicate the need for authors, peer reviewers, editors and readers to critically evaluate the content and quality of information presented in narrative reviews. We hope that this novel analysis prompts such action for future IPF review articles and that a similar introspective approach is used to assess the content and quality of narrative reviews in other areas of medicine.

Author contributions: C.J. Ryerson and K.M. Milne designed the study. K.M. Milne and C. Chan conducted the data acquisition. C.J. Ryerson and K.M. Milne performed the data analysis and produced the first draft of the manuscript. All authors provided critical input on study design, data analysis and manuscript preparation. C.J. Ryerson is the guarantor of this work.

Conflict of interest: K.M. Milne has nothing to disclose. C. Chan has nothing to disclose. J.H. Fisher has nothing to disclose. K. de Boer reports grants and personal fees from Boehringer Ingelheim, and personal fees from Hoffmann-La Roche and AstraZeneca, outside the submitted work. C.J. Ryerson reports grants and personal fees from Boehringer Ingelheim and Hoffmann-La Roche, outside the submitted work.

## References

- 1 Idiopathic pulmonary fibrosis: diagnosis and treatment. International consensus statement. American Thoracic Society (ATS), and the European Respiratory Society (ERS). *Am J Respir Crit Care Med* 2000; 161: 646–664.
- 2 Raghu G, Collard HR, Egan JJ, *et al.* An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med* 2011; 183: 788–824.
- 3 Raghu G, Rochwerg B, Zhang Y, *et al.* An official ATS/ERS/JRS/ALAT clinical practice guideline: treatment of idiopathic pulmonary fibrosis. An update of the 2011 clinical practice guideline. *Am J Respir Crit Care Med* 2015; 192: e3–e19.

- 4 The Idiopathic Pulmonary Fibrosis Clinical Research Network. Prednisone, azathioprine, and N-acetylcysteine for pulmonary fibrosis. *N Engl J Med* 2012; 366: 1968–1977.
- 5 King TE Jr, Bradford WZ, Castro-Bernardini S, et al. A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *N Engl J Med* 2014; 370: 2083–2092.
- 6 Noble PW, Albera C, Bradford WZ, et al. Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. *Lancet* 2011; 377: 1760–1769.
- 7 Richeldi L, du Bois RM, Raghu G, et al. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med* 2014; 370: 2071–2082.
- 8 Loke YK, Derry S. Does anybody read “evidence-based” articles? *BMC Med Res Methodol* 2003; 3: 14.
- 9 Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009; 339: b2535.
- 10 Charnock D, Shepperd S, Needham G, et al. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. *J Epidemiol Community Health* 1999; 53: 105–111.
- 11 Fisher JH, O’Connor D, Flexman AM, et al. Accuracy and reliability of internet resources for information on idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2016; 194: 218–225.
- 12 Demedts M, Behr J, Buhl R, et al. High-dose acetylcysteine in idiopathic pulmonary fibrosis. *N Engl J Med* 2005; 353: 2229–2242.
- 13 Global Initiative for Asthma (GINA). Global Strategy for Asthma Management and Prevention. 2018. Available from: <https://ginasthma.org/>
- 14 Mooney JJ, Collard HR. COUNTERPOINT: Should BAL be routinely performed in the diagnostic evaluation of idiopathic pulmonary fibrosis? No. *Chest* 2017; 152: 919–922.
- 15 Wells AU, Kokosi MA. POINT: Should BAL be routinely performed in the diagnostic evaluation of idiopathic pulmonary fibrosis? Yes. *Chest* 2017; 152: 917–919.