What forecasting the prevalence of inflammatory bowel disease may tell us about its evolution on a national scale

Mafalda Santiago, Fernando Magro (iD), Luís Correia, Francisco Portela, Paula Ministro, Paula Lago and Cláudia Camila Dias (iD); on behalf of GEDII (Portuguese IBD Group)

Abstract: Inflammatory bowel disease (IBD) is increasingly prevalent within western societies. Its complex and chronic facets in addition to its increasing prevalence place a great economic burden on our healthcare systems. Our aim was to estimate the national prevalence of IBD through predictive models. We used prevalence data which spans the years 2003–2007 to estimate prevalence until 2030 by means of four forecasting methods. Prevalence rates are estimated to be 4–6-times higher in 2030 when compared with 2003 with an average annual percent change of 5%. IBD is poised to have a substantial impact on healthcare systems in the near future, given its rapidly increasing prevalence. Forecasting methods will allow for a proactive stance on the development of health policies that will be needed to provide high quality and cost-effective care to these patients, while ensuring the economic viability of healthcare systems.

Keywords: forecasting, inflammatory bowel disease, predictive modeling, prevalence

Introduction

Inflammatory Bowel Disease (IBD) is a chronic immunologically mediated disorder of unknown etiology comprising Crohn's disease (CD) and ulcerative colitis (UC). It is estimated to affect up to 0.8% of the general population of western nations,[1] with accelerating incidence in newly industrialized countries whose societies have become more westernized due to rapid socioeconomic development.[2] Moreover, given its chronic nature, early onset, and low mortality, IBD prevalence has been increasing at a higher rate than incidence. This phenomenon is called compounding prevalence.[3–5]

In addition, IBD is associated with a considerable economic burden,[6] mostly due to the use of biologic therapy as well as the clinical and surgical hospitalizations that stem from the unpredictable relapsing and remitting course of the disease.[7] Consequently, healthcare systems and society will be progressively burdened by the increasing prevalence of IBD.[2,5]

Prediction modeling will allow for a proactive stance on the development of health policies that will be needed to tackle the impact of the increasing prevalence of IBD and the consequent challenges faced by healthcare systems.

In this short report, we follow up on the work of Azevedo and colleagues[8] in order to assess the evolution of IBD prevalence in Portugal by means of several forecasting methods.

Short report (methods + results)

Methods
For our dataset, we used the IBD prevalence estimates of Azevedo and colleagues[8] which span from 2003 to 2007. Prevalence data were then predicted from 2008 to 2030 by using four methods:

1. **Moving average method (MAM)**: prevalence evolution was assumed to be equal to the average of the previous 4 years.

Correspondence to:
Fernando Magro
Grupo de Estudo da Doença Inflamatória Intestinal (GEDII), Porto, Portugal

Centro Hospitalar de São João, Gastroenterology Department, Porto, Portugal

Department of Biomedicine, Unit of Pharmacology and Therapeutics, Faculty of Medicine, University of Porto, Alameda Prof. Hernâni Monteiro 420-319 Porto, Portugal
fm@med.up.pt

Mafalda Santiago
CINTESIS – Center for Health Technology and Services Research, Porto, Portugal

Grupo de Estudo da Doença Inflamatória Intestinal (GEDII), Porto, Portugal

Luís Correia
Grupo de Estudo da Doença Inflamatória Intestinal (GEDII), Porto, Portugal

Centro Hospitalar Lisboa Norte, Hospital Santa Maria, Gastroenterology Department, Lisboa, Portugal

Francisco Portela
Grupo de Estudo da Doença Inflamatória Intestinal (GEDII), Porto, Portugal

Centro Hospitalar Universitário de Coimbra, Gastroenterology Department, Coimbra, Portugal

Paula Ministro
Grupo de Estudo da Doença Inflamatória Intestinal (GEDII), Porto, Portugal

Centro Hospitalar Tondela e Viseu, Gastroenterology Department, Viseu, Portugal

Paula Lago
Grupo de Estudo da Doença Inflamatória Intestinal (GEDII), Porto, Portugal

Centro Hospitalar do Porto, Hospital Geral Santo António, Gastroenterology Department, Porto, Portugal

Cláudia Camila Dias
CINTESIS – Center for Health Technology and Services Research, Porto, Portugal

Department of Community Medicine, Information and Health Decision Sciences, Faculty of Medicine of the University of Porto, Portugal

$$\bar{t}_{SM} = \frac{t_M + t_{M-1} + \ldots + t_{M-(n-1)}}{n} = \frac{1}{n}\sum_{i=0}^{n-1} t_{M-i}$$

2. **Drift method (DM):** in this method, the amount of change over time (also called drift) is the average seen in the historical data. Hence, the forecast for time $T+h$ is

$$\hat{y}_{T+h|T} = y_T + \frac{h}{T-1}\sum_{t=2}^{T}(y_t - y_{t-1})$$

$$= y_T + h\left(\frac{y_T - y_1}{T-1}\right)$$

, which is equivalent to extrapolating a line drawn between the first and last data points.

3. **Error, Trend and Seasonality method (ETSM)**: exponential smoothing forecasting methods are similar in that a prediction is a weighted sum of past observations, however, there is an exponentially decreasing weight for past observations. There are several exponential smoothing methods, so the ETS framework provides an automatic way of selecting the best method of exponential smoothing to be implemented. In our case, it was the Holts linear method with additive errors

$$\hat{y}_{t+h|t} = l_t$$

$$l_t = ay_t + (1-a)(l_{t-1} + b_{t-1})$$

$$b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1}, \text{ where}$$

α and β^* are two smoothing parameters ranging from 0 to 1; l_t level is the weighted average between y_t one-step ahead forecast time t, $(l_t + b_{t-1} = \hat{y}_{t|t-1})$; b_t slope is the weighted average of $(l_t - l_{t-1})$ and b_{t-1} is the current and previous estimate of the slope.

4. **Theta decomposition method of forecasting (TM):** simple exponential smoothing (SES) with drift.[9] An alternative forecasting method to that which was described by Assimakopoulos and Nikolopoulos.[10]

$$\hat{X}_n(h) = \bar{X}_n(h) + \frac{1}{2}\hat{b}_0\left(h - 1 + \frac{1}{a} - \frac{(1-a)^n}{a}\right),$$

where

$\bar{X}_n(h)$ is the SES forecast of the series $\{X_t\}$; $\hat{b}_0$ is the parameter for linear trend

fitted to the original series $\{X_t\}$ and a is the smoothing parameter for the SES.

These naïve forecasting methods (DM, ETSM and TM) were chosen due to the size and quality of our dataset and implemented using the forecast package in R.[11,12] Accuracy tests were performed in order to evaluate the fitting of the data for all methods apart from the MA method which is a simple moving average calculation (Supplementary Table 1).

Prediction intervals (PIs; 80% and 95%) were used in the visual display of the DM, ETSM and TM methods. Confidence intervals (CIs; 80% and 95%) were used for the visual display of the MAM.

Results

Prediction methods
The four prediction methods yielded similar prevalence results throughout the forecasted period with the exception of the TM method, which provided more conservative estimates (Table 1).

To evaluate the fitting of the forecasts, we performed accuracy tests. The DM presented the smallest error values regardless of which accuracy measure was used (Supplementary Table 1), and therefore may be considered the most accurate of those implemented and the most similar to the simple MAM.

Forecasted IBD prevalence in Portugal
In 2019, the calculated prevalence of IBD in Portugal is: MAM: 319 per 100,000 (95% CI: 286–351); DM: 326 per 100,000 (95% PI: 241–411); ETSM: 302 per 100,000 (95% PI: 254–350); and TM: 325 per 100,000 people (95% PI: 145–334). In 2030, the forecasted prevalence ranges from 325 per 100,000 (TM 95% PI: 194–456) to 491 per 100,000 people (DM: 95% PI: 332–650).

When comparing data from 2003 to the forecasted prevalence of 2030, we obtained prevalence estimates 4–6-fold higher (Table 1). In addition, the average annual percent change of the forecasted prevalence rate was approximately 4–6%, regardless of the methods implemented. This meaning that the prevalence of IBD is forecasted to climb 4–6% per year in Portugal (Table 1).

Table 1. Predicted prevalence rates between 2003 and 2030 through four forecasting methods.

Year	MAM	CI (80%)	CI (95%)	DM	PI (80%)	PI (95%)	ETSM	PI (80%)	PI (95%)	TM	PI (80%)	PI (95%)
2003	86			86			86			86		
2004	99			99			99			99		
2005	121			121			121			121		
2006	135			135			135			135		
2007	146			146			146			146		
2008	161	148.0 174.0	141.1 180.9	161	154.1 167.9	150.4 171.6	159	149.8 168.2	145.0 173.0	154	136.0 171.6	126.5 181.1
2009	177	162.9 190.1	155.7 197.3	176	164.4 187.6	158.3 193.7	172	159.1 184.9	152.2 191.8	162	136.4 186.8	123.0 200.2
2010	190	175.8 204.9	168.1 212.7	191	174.9 207.1	166.4 215.6	185	169.1 200.9	160.8 209.2	169	138.5 200.3	122.1 216.7
2011	204	188.8 219.6	180.7 227.8	206	185.5 226.5	174.6 237.4	198	179.7 216.3	170.0 226.0	177	141.5 212.9	122.6 231.8
2012	219	202.6 235.0	194.0 243.5	221	196.0 246.0	182.8 259.2	211	190.5 231.5	179.7 242.3	185	145.1 224.9	124.0 246.0
2013	233	216.2 250.2	207.2 259.2	236	206.6 265.4	191.1 280.9	224	201.6 246.4	189.7 258.3	193	149.1 236.5	125.9 259.7
2014	247	229.6 265.2	220.2 274.6	251	217.2 284.8	199.4 302.6	237	212.8 261.2	200.0 274.0	201	153.4 247.8	128.4 272.8
2015	262	243.1 280.2	233.3 290.0	266	227.8 304.2	207.6 324.4	250	224.1 275.9	210.4 289.6	208	157.9 258.9	131.2 285.6
2016	276	256.8 295.2	246.6 305.4	281	238.4 323.6	215.9 346.1	263	235.5 290.5	221.0 305.0	216	162.7 269.7	134.3 298.1
2017	290	270.4 310.3	259.8 320.8	296	249.1 342.9	224.2 367.8	276	247.1 304.9	231.7 320.3	224	167.6 280.4	137.7 310.3
2018	305	284.0 325.2	273.1 336.1	311	259.7 362.3	232.5 389.5	289	258.6 319.4	242.6 335.4	232	172.6 291.0	141.3 322.3
2019	319	297.6 340.2	286.4 351.4	326	270.3 381.7	240.8 411.2	302	270.3 333.7	253.5 350.5	240	177.8 301.4	145.0 334.2
2020	333	311.3 355.1	299.7 366.7	341	280.9 401.1	249.1 432.9	315	282.0 348.0	264.5 365.5	247	183.0 311.8	149.0 345.8
2021	348	325.0 370.0	313.0 382.0	356	291.5 420.5	257.4 454.6	328	293.8 362.2	275.6 380.4	255	188.4 322.0	153.1 357.3
2022	362	338.7 384.9	326.4 397.2	371	302.1 439.9	265.7 476.3	341	305.6 376.4	286.8 395.2	263	193.9 332.1	157.3 368.7
2023	376	352.4 399.8	339.8 412.4	386	312.8 459.2	274.0 498.0	354	317.4 390.6	298.0 410.0	271	199.4 342.2	161.6 380.0
2024	390	366.1 414.7	353.2 427.6	401	323.4 478.6	282.3 519.7	367	329.3 404.7	309.3 424.7	279	205.0 352.2	166.0 391.2
2025	405	379.8 429.6	366.7 442.7	416	334.0 498.0	290.6 541.4	380	341.2 418.8	320.6 439.4	286	210.7 362.1	170.6 402.2
2026	419	393.6 444.4	380.1 457.9	431	344.6 517.4	298.9 563.1	393	353.1 432.9	332.0 454.0	294	216.4 372.0	175.2 413.2
2027	433	407.3 459.3	393.6 473.0	446	355.2 536.8	307.2 584.8	406	365.1 446.9	343.4 468.6	302	222.2 381.8	179.9 424.1
2028	448	421.1 474.1	407.1 488.1	461	365.8 556.2	315.5 606.5	419	377.1 460.9	354.9 483.1	310	228.0 391.6	184.7 434.9
2029	462	434.9 488.9	420.6 503.2	476	376.5 575.5	323.8 628.2	432	389.1 474.9	366.3 497.7	318	233.9 401.3	189.6 445.6
2030	476	448.7 503.7	434.1 518.3	491	387.1 594.9	332.1 649.9	445	401.1 488.9	377.9 512.1	325	239.8 411.0	194.5 456.3

CI, confidence interval; DM, drift method; ETSM, ETS method; MA, moving average method; PI, prediction interval; TM, theta method.
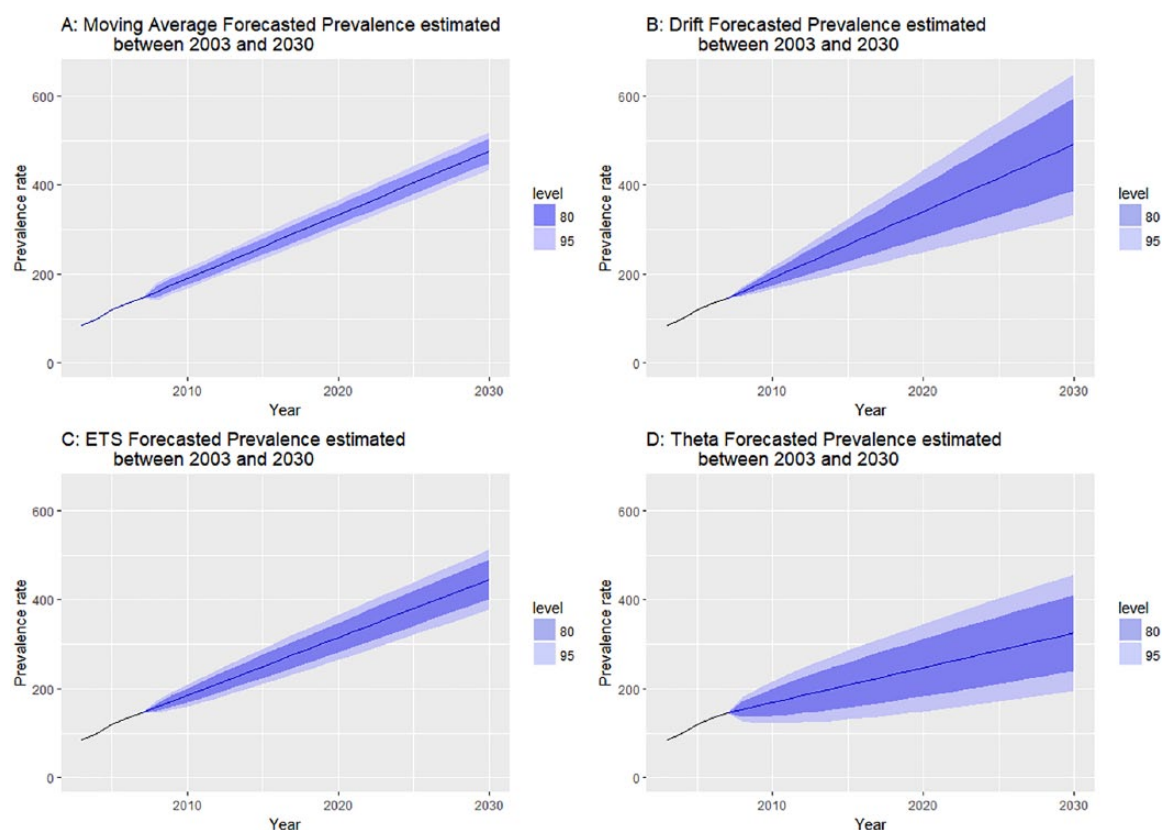
Figure 1. Forecasted prevalence rates of IBD between the years 2003 and 2030. Shaded areas are either confidence intervals (a) or prediction intervals (b, c, d).
IBD, inflammatory bowel disease.

Figure 1 presents the annual forecasted prevalence of IBD for each method, with associated 80% and 95% PI or CI.

Discussion

Our analysis, based on the application of four simple forecasting methods pertaining to previously published data, allowed the estimation of past, present and future prevalence of IBD in a southern European country.

In 2003, approximately 0.1% of the Portuguese population was diagnosed with IBD. According to our estimations, this figure is forecasted to rise to 0.22–0.33% in the past year of 2016, reaching 0.24–0.33% in 2019 and 0.32–0.49% in 2030. Consequently, Portugal may be considered part of the group of European countries with the highest IBD prevalence estimates. Since in Europe, the highest reported prevalence values correspondent to the period of 1990 to 2016, were 505 per 100,000 people for UC and 322 per 100,000 people for CD.[2]

As previously mentioned, the prevalence of IBD is estimated to climb approximately 4–6% per year in Portugal. This is due to the imbalance between the incidence and mortality of IBD as a chronic and incurable disorder.[13]

Moreover, these estimations are higher than what was predicted by Coward and colleagues[14] for Canada (2.86%), which may be explained by the different methodologies applied to each study. Whereas Coward and colleagues[14] study analyzed larger datasets presenting a more robust implementation of forecasting methods, our work has a more comparative nature due to the limited available data.

Since the increase in IBD prevalence appears to be inevitable, it is of major importance for health policies to be developed and healthcare systems

adapted to the growing burden of IBD and the resulting economic impact.[5,6,15]

Additionally, more studies are needed in order to reach a better understanding of the actual prevalence of IBD in Portugal, since Azevedo and colleagues[8] used a pharmaco-epidemiological approach based of the use of intestinal anti-inflammatory drugs to reach the original prevalence estimates.

Regarding the forecasting analysis, our estimations were based on the prevalence estimates reported by Azevedo and colleagues[8] which comprise only a 5-year period. Although accuracy tests were performed, these results must be taken with caution as we must acknowledge the severe limitations of conducting forecast estimations on a small dataset. This was the rationale for the implementation of several simple forecasting methods, in order to allow comparisons among the results obtained.

The DM, which has shown better accuracy, estimates the slope of the original sample, draws a line and extrapolates it into the future. The MAM only considers a predetermined number of past data points, four in our case. These two methods yielded the most comparable results due to the similarity between the size of our sample and the number of data points considered for the calculation of MAM. As for the ETSM, it presents two smoothing parameters, alpha and beta with values ranging from 0 to 1. The first when close to 1 means estimates are based on recent observations while a small beta value, which represents the slope of the time series, implies a slightly unchanging trend.[16] This is representative of our forecasted scenario regarding the results by the ETSM. Hence, the smaller estimates and respective PIs (Table 1). As previously mentioned, the TM can be described as a special case of SES with drift that is equal to half the slope of a straight line fitted to the data.[9] Therefore, presenting the more conservative prevalence prediction (Table 1).

Additionally, we tested the aforementioned models on a dataset of the prevalence of CD from the study of Yao and colleagues.[17] This was a nationwide study which covered the period of 1986–1998. We compared the results of the forecasting methods to the work of Asakura and colleagues, to our knowledge the most recent nationwide prevalence study of IBD in Japan. Asakura and colleagues reported a CD prevalence of 21.2 per 100,000 persons in the year 2005.[18]

As seen in Supplementary Figure 1, our CD prevalence forecasts for the year 2005 were similar to what was estimated by Asakura and colleagues. Moreover, according to the accuracy tests performed, the model showing the lowest error values was the ETSM which presents the closest estimation of that of Asakura and colleagues.[18]

Overall, these results show the importance of performing future prevalence and forecasting studies and to act on the impending burden of IBD, given the escalation of IBD prevalence on such a likely high rate, as estimated here.

ORCID iDs
Fernando Magro https://orcid.org/0000-0003-2634-9668
Cláudia Camila Dias https://orcid.org/0000-0001-9356-3272

Supplemental material
Supplemental material for this article is available online.

References

1. Molodecky NA, Soon IS, Rabi DM, *et al*. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology* 2012; 142: 46–54.e42.

2. Ng SC, Shi HY, Hamidi N, *et al*. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet* 2017; 390: 2769–2778.

3. Vatn MH. Natural history and complications of IBD. *Curr Gastroenterol Rep* 2009; 11: 481–487.

4. Cosnes J, Cattan S, Blain A, *et al*. Long-term evolution of disease behavior of Crohn's disease. 2002; 8: 244–250.

5. Kaplan GG. The global burden of IBD: from 2015 to 2025. *Nat Rev Gastroenterol Hepatol* 2015; 12: 720–727.

6. Stone CD. The economic burden of inflammatory bowel disease: clear problem, unclear solution. *Dig Dis Sci* 2012; 57: 3042–3044.

7. Van Der Valk ME, Mangen MJJ, Leenders M, *et al*. Healthcare costs of inflammatory bowel disease have shifted from hospitalisation and surgery towards anti-TNFα therapy: results from the COIN study. *Gut* 2014; 63: 72–79.

8. Azevedo LF, Magro F, Portela F, *et al*. Estimating the prevalence of inflammatory bowel disease in Portugal using a pharmaco-epidemiological approach. *Pharmacoepidemiol Drug Saf* 2010; 19: 499–510.

9. Hyndman RJ and Billah B. Unmasking the theta method. *Int J Forecast* 2003; 19: 287–290.

10. Assimakopoulos V and Nikolopoulos K. The theta model: a decomposition approach to forecasting. *Int J Forecast* 2000; 16: 521–530.

11. Hyndman RJ and Athanasopoulos G. *Forecasting: principles and practice*. OTexts.com. Heathmont, Victoria, 2014.

12. Hyndman R, Athanasopoulos G, Bergmeir C, *et al. Forecast: forecasting functions for time series and linear models*. R Package version 8.5e, 2019.

13. Macdonald G (2008) Harrison's internal medicine, 17th edition. - by AS Fauci, DL Kasper, DL Longo, E Braunwald, SL Hauser, JL Jameson and J Loscalzo. *Internal Med J* 2008; 38: 932–932.

14. Coward S, Clement F, Benchimol EI, *et al*. Past and future burden of inflammatory bowel diseases based on modeling of population-based data. *Gastroenterology* 2019; 156: 1345–1353, e4.

15. Pillai N, Dusheiko M and Burnand B, *et al*. A systematic review of cost-effectiveness studies comparing conventional, biological and surgical interventions for inflammatory bowel disease. *PLoS One* 2017; 12: p.e0185500.

16. Hyndman JR, Anne KB, Ralph SD, *et al*. A state space framework for automatic forecasting using exponential smoothing methods. *Int J Forecast* 2002; 18: 439–454.

17. Yao T, Matsui T and Hiwatashi N. Crohn's disease in Japan: diagnostic criteria and epidemiology. *Dis Colon Rectum* 2000; 43(Suppl. 10): S85–S93.

18. Asakura K, Nishiwaki Y, Inoue N, *et al*. Prevalence of ulcerative colitis and Crohn's disease in Japan. *J Gastroenterol* 2009; 44: 659–665.