Aetiology of ulcerative colitis

II A new hypothesis

P. R. J. BURCH, F. T. de DOMBAL, AND G. WATKINSON

None of the hypotheses reviewed in our previous paper tackled the pathogenetic issues posed by the sex and age distributions of ulcerative colitis. In this paper we show that the age patterns of seven large clinical series, from three different continents, display some remarkably consistent features. From a detailed analysis of the epidemiological statistics, and from clinical, genetic, and serological evidence, we are able to put forward a new aetiological theory.

We shall argue that ulcerative colitis develops only in those people who have a genetic predisposition to the disease. Evans and Acheson (1965) have previously distinguished between early- and late-onset cases, and our studies, both of the age distributions, and of the clinical course of the disease in early- and late-onset patients, confirm the idea that at least two distinctive groups are predisposed.

Given the diathesis, the initiation of ulcerative colitis appears to depend on random intrinsic events, of which the average rate of occurrence is effectively invariant at different times and in widely separated environments. Following Burnet (1959, 1965), we suggest these intrinsic random events correspond to somatic gene mutations in mesenchymal stem cells. Ulcerative colitis appears to belong to the general class of disease widely known as ‘disturbed-tolerance autoimmunity’. However, many reasons can be adduced for rejecting the notion that immunoglobulin autoantibodies constitute the primary causal agents in idiopathic ‘autoimmunity’ (Fahey and Goodman, 1964; Samloff and Barnett, 1965; Bernier and Hines, 1967; Burch, Rowell, and Burwell, 1968). From our serological findings (de Dombal, 1967a, b), we conclude that if the primary ‘autoantibodies’ are humoral, they are more likely to be found in the α₂-globulin, than in the β- or γ-globulin electrophoretic fractions. Accordingly, we suggest that ‘autoaggressive’ is, perhaps, a less misleading description than ‘autoimmune’, for it conveys no prejudicial overtones regarding the identity of the cell system(s) producing the primary pathogenic agents.

SEX-SPECIFIC AND AGE-SPECIFIC ONSET RATES

Aetiological theories have been tested for their ability to account for the sex and age distributions of diseases that are widely believed to be ‘autoimmune’, such as chronic discoid and systemic lupus erythematosus, inflammatory polyarthritics, and myasthenia gravis (Burch, 1966b, 1968; Burch and Rowell, 1965a, 1968). Only Burnet’s ‘forbidden-clone’ theory of disturbed-tolerance autoimmunity (Burnet, 1959, 1965; Burnet and Holmes, 1965) provides a convincing interpretation of the age patterns of such diseases.

Briefly, this theory states: (1) predisposition to ‘autoimmune’ disease is genetically determined; (2) the growth of forbidden clones is initiated by the random occurrence of specific somatic mutations in mesenchymal stem cells; (3) forbidden clones synthesize humoral or cellular ‘autoantibodies’ that attack tissues bearing complementary antigens; and (4) a latent period necessarily intervenes between the last initiating somatic mutation and the first manifestation of the disease.

We find the duration of the latent period often depends, to some extent, on environmental factors (Burch, 1966b). Granted certain provisos, the age patterns of all spontaneous disturbed-tolerance autoimmune diseases—or better, autoaggressive diseases—are expected to conform to certain simple mathematical rules (Burch, 1966a, b). For example, we should be able to describe age-specific initiation rates by stochastic equations such as 1 and 2 below.

Ideally, statistics for age analysis would comprise every confirmed case, by sex and age at first onset, in several very large, geographically separated, and precisely defined populations, at different secular times. Needless to say, such ideal data do not exist, but provided there is no systematic age or sex bias in the admission of patients to a clinical series, then no distortion of the age pattern will arise from selection factors. Indeed, clinical series and population studies lead to the same general conclusions (Burch, 1968).

We have analysed data from seven clinical series.
of ulcerative colitis patients, including our own, and the results are presented graphically in Figures 1 to 7. In view of the hazards of clinical material, the consistencies in these extensive data are remarkable. From the bimodal features of some distributions, and the late bulge in others, it appears that two distinctive groups (at least) are at risk with respect to ulcerative colitis. The age distribution for the larger, early-onset group, is described by the Weibull equation (Burch, 1966a):

\[
dP/dt = 2k_1S_1t \exp (-k_1t^2)
\]

where \(k_1\) has the same value (1·5 \times 10^{-3}yr^{-3}) for every series studied \(dP/dt\) defines the age-specific initiation rate of the disease at age \(t\) years. This is given in absolute units for the Norwegian series (Fig. 2) but in relative terms for the others. The definition of the constant \(k_1\) depends on the biological model, but it is related to the kinetics of initiating events. On the simplest and most plausible model, \(k_1 = L_1m_1^2\), where \(L_1\) represents the number of cells at mutational risk, and \(m_1\) describes the average rate of somatic gene mutation for each of two genes at mutational risk, in each of the \(L_1\) cells. We require \(m_1t \ll 1\) for all \(t\) of interest (for details, see Burch 1966a). \(S_1\) defines the proportion of the population (specific for sex) that is predisposed at birth to the early-onset form of the disease.

For the late-onset group (initiation mode at 50 years) the age pattern is described by the following Weibull equation where \(k_2\) has the value 5·5 \times 10^{-11}yr^{-6}:

\[
dP/dt = 6k_2S_2t^6 \exp (-k_2t^4)
\]

The symbols in this equation have analogous definitions to those in (1), except that \(k_2 = L_2m_2^4\).

TWO PREDISPOSED GROUPS: CLINICAL EVIDENCE

Evans and Acheson (1965) have drawn attention to the bimodality of age-specific onset rates in the Oxford series, and they have suspected the existence of two distinct groups of patients. Diarrhoea was found to be a more common, and rectal bleeding was a less common presenting symptom in the late- than in the early-onset cases.

We have examined our records for differences in the clinical pattern related to age at onset. When onset occurs above 60 years of age, a relatively low relapse rate is found (Table I). This is not due (see Table II) to the fact that these patients are simply over 60 (group A) or have been closely followed (group B), or presented soon after the onset of colitis (group C). Hence the clinical evidence supports the indications given by the various age patterns (Figs. 1 to 7) for the existence of two distinct groups of colitis patients. Cases belonging to the early-onset group show a modal initiation age at 18 years; the late-onset group has a modal initiation age at 50 years.

IMPLICATIONS OF AGE DISTRIBUTION

We shall not enter here into elaborate discussions of the mathematics and of the biological implications of the age patterns of autoaggressive diseases,
Aetiology of ulcerative colitis

**FIG. 2.** Absolute initiation-rates (dP/dt), combined for sex, as a function of estimated age at initiation (t) derived from the Norwegian hospital series of Ustvedt (1958). The long latent period correction (18 years) for first admission to hospital (not onset) cannot be applied to the first two decades. For the second decade, the applied correction is 13 years. The proportion of the Norwegian population (S1) (males and females) predisposed to early-onset ulcerative colitis, and admitted to hospital, is about $6 \times 10^{-4}$; the value of $S_2$, or late-onset ulcerative colitis is about $3 \times 10^{-4}$. More females are predisposed than males in both groups, but numbers are too small for reliable assessment of the sex ratios. The overall ratio, $S_1/S_2$, is 2:1. Otherwise as for Figure 1. Early-onset cases (a) are fitted to the equation: $\frac{dp}{dt} \propto t \exp(-k_1t^2)$ and late-onset cases (b) to: $\frac{dp}{dt} \propto t^6 \exp(-k_4t^4)$.

**FIG. 3.** Relative age-specific initiation rates (dP/dt) combined for sex as a function of estimated age at initiation derived from the Lahey Clinic series of Fernandez-Herlihy (1959). Latent period correction to age at onset (all decades except first) is nine years. The overall sex ratio for this series ($S_F/S_M$) is 1:1. The ratio $S_1/S_2$ of early- to late-onset cases is about 15. Early-onset cases (a) are fitted to the equation: $\frac{dp}{dt} \propto t \exp(-k_1t^2)$ and late-onset cases (b) to: $\frac{dp}{dt} \propto t^6 \exp(-k_4t^4)$.

**FIG. 4.** Relative age-specific initiation rates (dP/dt), sexes combined, as a function of estimated age at initiation (t) for the New Zealand series of Wigley and MacLaurin (1962). Latent period correction (initiation to onset) is 17 years for all decades except the first, where 4.5 years is assumed. The sex-ratio (overall) is approximately unity, although total numbers (96 males and 93 females) are too small for accurate definition. The ratio of early- to late-onset groups, $S_1/S_2$, is approximately 4. Early-onset cases (a) are fitted to the equation: $\frac{dp}{dt} \propto t \exp(-k_1t^2)$ and late-onset cases (b) to: $\frac{dp}{dt} \propto t^6 \exp(-k_4t^4)$.

**FIG. 5.** Relative sex-specific and age-specific initiation rates (dP/dt) as a function of estimated age at initiation (t) for the Oxford series of Edwards and Truelove (1963). Latent period correction (initiation to onset) is 10 years for all decades except the first, where four years is assumed. For the early-onset group, the sex ratio, $S_{1,F}/S_{1,M}$, is 1:4:1; for the late-onset group, $S_{2,F}/S_{2,M}$ is about 1:3. The ratio: $S_{1,M}/S_{2,M}$ is about 3:7. Early-onset cases (a) are fitted to the equation: $\frac{dp}{dt} \propto t \exp(-k_1t^2)$ and late-onset cases (b) to: $\frac{dp}{dt} \propto t^6 \exp(-k_4t^4)$. 
FIG. 6. Relative sex-specific and age-specific initiation rates (dP/dt) as a function of estimated age at initiation (t) for the London series of MacDougall (1964). Latent period correction (initiation to hospital registration) is 15 years for all decades except the first two. A correction of 14 years has been applied to the data for the second decade. For the early-onset group, the sex ratio S1,F/S1,M is 1-2; for the late-onset group S2,F/S2,M \( \simeq 2 \). For the early-onset group (a) are fitted to the equation: 
\[
\frac{dP}{dt} \propto t \exp \left( -k_1 t^2 \right) \quad \text{and late-onset cases (b) to: } \quad \frac{dP}{dt} \propto t^6 \exp \left( -k_4 t^6 \right).
\]

TABLE I

<table>
<thead>
<tr>
<th>Age at Onset (year)</th>
<th>Total Observed Patient Years</th>
<th>Patient Years with at Least One Attack (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>60 and over</td>
<td>99</td>
<td>26</td>
</tr>
<tr>
<td>50 to 59</td>
<td>71</td>
<td>35</td>
</tr>
<tr>
<td>Under 50</td>
<td>69</td>
<td>48</td>
</tr>
</tbody>
</table>

*One patient year is one complete year of follow-up of an individual patient.*

because these have been presented in detail elsewhere (Burch, 1966a, b, 1968; Burch and Rowell, 1965a, b, 1968). Suffice it to say, that the age patterns of colitis in these seven series from three different continents show some remarkably consistent features which are also displayed by so-called disturbed-tolerance ‘autoimmune’, or autoaggressive diseases. Values of \( k_1 \) and \( k_2 \) stay very effectively the same from series to series. However, sex ratios \( (S_{1,M}/S_{1,F}; S_{2,M}/S_{2,F}) \), and the proportions \( (S_1/S_2) \) of early- to late-onset groups differ from series to series, and these features reflect differences in population gene frequencies between and within countries.

If the environment fails to affect the rate of initiation of a disease (as characterized by the constants \( k_1 \) and \( k_2 \)), then either the disease must be caused by extrinsic factors whose impact stays

---

**TABLE II**

<table>
<thead>
<tr>
<th>Patient Category</th>
<th>No. of Patient Years with at Least One Attack (%)</th>
<th>Patient Years with at Least One Attack (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Patient years in patients over 60 years</td>
<td>Onset 60 years or over</td>
</tr>
<tr>
<td></td>
<td>(from Table I)</td>
<td>Onset 0 to 59 years</td>
</tr>
<tr>
<td></td>
<td>All patient years in patients systematically</td>
<td>140</td>
</tr>
<tr>
<td></td>
<td>followed from first attack onwards</td>
<td></td>
</tr>
<tr>
<td></td>
<td>All patient years within 10 years</td>
<td>974</td>
</tr>
</tbody>
</table>

---

FIG. 7. Relative sex-specific and age-specific initiation rates (dP/dt) as a function of estimated age at initiation (t) for the Leeds General Infirmary series (Watts, de Dombal, Watkinson, and Galigher, 1965). Latent period correction (initiation to onset) for all decades except the first is 10 years; for the first decade four years is assumed. The sex ratio, \( S_{1,F}/S_{1,M} \), for the early-onset group is 1-4; for the late-onset group \( S_{2,F}/S_{2,M} \) is about 0-9. The ratio \( S_{1,F}/S_{2,M} \) is about 3-4. Early-onset cases (a) are fitted to the equation: 
\[
\frac{dP}{dt} \propto t \exp \left( -k_1 t^2 \right) \quad \text{and late-onset cases (b) to: } \quad \frac{dP}{dt} \propto t^6 \exp \left( -k_4 t^6 \right).
\]
early-onset ulcerative colitis is X-linked factors to about an induction mechanism genetic geography with (Burch, mellitus expression the sex dominate. Then terms, dominant give a of ulcerative Olsen, and and Si,S1,FIS1, Clinic series the 1962), interesting of ulcerative colitis and others genetically based diseases, including Crohn's disease, spondylitis, ankylosing spondylitis, uveitis, chronic hepatic cirrhosis, eczema, and hay fever (Tumen, Monoghan, and Jobb, 1947; Pollard and Block, 1948; Comfort, Bargen, and Morlock, 1938; Willcox and Isselbacher, 1961; Holdsworth, Hall, Dawson, and Sherlock, 1965; Wright, Lumsden, Luntz, Sevel, and Truelove, 1965; Wright and Watkinson, 1965a, b; Binder et al, 1966; Billson, de Dombal, Watkinson, and Goligher, 1967; Korelitz and Coles, 1967; Hammer, Ashurst, and Naish, 1968). If one or more of the genes predisposing to ulcerative colitis also predisposes to these other diseases, then the positive associations between them can be explained. Genes with pleiotropic effects are well known. In the present context such genes are likely to encode for antigenic determinants (histocompatibility antigens or 'tissue coding factors') that are common to the colonic mucosa and to the other tissues attacked by other forbidden clones in genetically associated diseases (Burch, 1968).

To summarize, predisposition to the early-onset form of ulcerative colitis is associated with one genotype, whereas the late-onset form is confined to a different genotype. (The two genotypes may well have certain alleles in common, as in the example of psoriasis (Burch and Rowell, 1965b)). Predisposition to both early- and late-onset forms of the disorder is polygenic, and positive associations with other genetically based diseases depend on the pleiotropic effects of one or more of the predisposing genes.

SEROLOGICAL EVIDENCE:
THE ROLE OF ALPHA-2-GLOBULINS

It is widely suspected (see Introduction) that immunoglobulin autoantibodies cannot be the primary pathogenic agents in autoaggressive ('autoimmune') disease. One of us has proposed that when the target tissue of the autoaggressive attack lies behind a blood-tissue barrier the primary pathogen usually migrates on electrophoresis with the a2-globulin serum protein fraction (Burch and Burwell, 1965;
Burch and Rowell, 1965b). If this view is applicable to autoaggressive attacks on the epithelial lining of the colon, increases in the level of the $\alpha_2$-globulin fraction would be expected to precede acute attacks of colitis, and remissions should be preceded by a fall in $\alpha_2$-globulin levels. The interval between the first increase in pathogenic $\alpha_2$-globulins and the onset of symptoms will depend on the extent and severity of the autoaggressive attack. Such considerations raise the prior question, Why should a disease such as ulcerative colitis show a fluctuating course? Burnet (1965) has argued that the activity of a forbidden clone depends on the outcome of two opposed tendencies: (1) the proliferative propensity of the forbidden clone itself, and (2) the efficiency of an endogenous defence or surveillance mechanism directed against the forbidden clone. Serological evidence indicates that the endogenous defence mechanism relies on immunoglobulin antibodies complementary to the mutant ("not-sell") cells of the forbidden clone (Burch and Rowell, 1965; Burch and Rowell, 1965a, b). Hence, competition for defence resources, from, for example, certain infective agents, can precipitate and exacerbate the latent disease condition. Also, factors that reduce the synthesis of defence immunoglobulins, either directly or indirectly through hormonal action, disturb the precarious balance between defence and attack. Acute mental stress may produce exacerbations in this way.

Our serological evidence agrees with this theory of endogenous defence (de Dombal, 1967a, b). Outpatients in remission who were free from infectious and incidental diseases such as influenza and bronchitis were chosen for study. Fifteen patients with levels of $\alpha_2$-globulin above the normal range suffered an acute attack within three months, while 35 patients with normal or diminished $\alpha_2$-globulin levels suffered no such attacks (de Dombal, 1967b). When at admission to hospital during an acute attack the patient's $\gamma$-globulin levels were raised, and/or rising, remission occurred. However, when the level of $\gamma$-globulins was low and falling, the patient invariably had to proceed to surgery (de Dombal, 1967b). These findings powerfully support the view that $\gamma$-globulins can play a defensive role.

Although it is well established that lymphoid and plasma cells synthesize IgG immunoglobulins, the type of cell synthesizing the (hypothetical) primary $\alpha_2$-globulin autoantibodies remains unknown. Similarly, the specific $\alpha_2$-globulin fraction(s) showing an increase before acute attacks has (have) yet to be determined. Studies on the concentrations of serum proteins taken from the inferior mesenteric artery and vein of patients with severe attacks of ulcerative colitis have shown that $\alpha_2$-globulins are synthesized locally within the colon (de Dombal, 1967a). From the literature of related autoaggressive diseases, it has been tentatively suggested that mast cells might be involved, directly or indirectly, in the synthesis of primary $\alpha_2$-globulin 'autoantibodies' (Burch and Burwell, 1965; Burch, 1968).

**The Role of Mast Cells**

If the primary 'autoantibodies' in ulcerative colitis are $\alpha_2$-globulins and if they are synthesized by mast cells, then because these proteins are produced in the colon (de Dombal 1967a) we might expect to find raised levels of mast cells in the colon of patients with ulcerative colitis. Histopathological observations confirm this expectation (McGovern and Archer, 1957; Priest, Rebuck and Havey, 1960; McAuley and Sommers, 1961; Rebuck, Hodson, Priest, and Barth, 1963; Rupe, 1963; Bercovitz and Sommers, 1966). Eosinophils and plasma cells are also unusually abundant, and the latter may well participate in a classical immune response to antigenic debris released during the primary autoaggressive attack on the colonic mucosa.

The involvement of mast cells in the histopathology, and the positive genetic associations between colitis and allergic conditions such as hay fever, provide a link between ulcerative colitis and immediate hypersensitivity. Diseases of this general class may cover a wide spectrum, to include those which at one extreme are manifested only when the appropriate environmental allergen is present, to those at the opposite extreme in which environmental factors play no essential pathogenetic role.

**Conclusions**

We can summarize our own aetiological proposals as follows:

1. Ulcerative colitis cannot be due primarily to extrinsic factors such as foreign antigens, milk, or microorganisms because these factors vary greatly from place to place, and from time to time, whereas the constants $k_1$ and $k_2$, describing the kinetics of the age distributions, do not.

2. Two groups (at least) are predisposed to ulcerative colitis. Predisposition is polygenic for each group.

3. Random events initiate the disease process, and following Burnet (1959, 1965) we believe that these are likely to be a special form of somatic gene mutation in mesenchymal stem cells. Probably two mutations (in a single stem cell) initiate the disease in a predisposed person belonging to the early-onset group. In a person belonging to the late-onset group, probably six such events initiate the disease process.
Aetiology of ulcerative colitis

4 The mutant mesenchymal stem cell divides to propagate a forbidden clone of cells synthesizing the primary 'autoantibodies'. After a latent period, typically of about nine to 12 years (average), depending to some extent on the environment, the patient experiences the first attack of colitis.

5 Serological investigations (de Dombal, 1967a, b) indicate that the primary 'autoantibodies' do not migrate on electrophoresis with the β- or γ-globulin serum protein fractions. If the primary 'autoantibodies' are humoral they are more likely to be found in the α-2-globulin fraction (de Dombal, 1967b).

6 Certain extrinsic factors, and severe mental stress, probably precipitate and exacerbate colitis by disturbing an endogenous defence mechanism directed against the forbidden clone.

SUMMARY

Mathematical analysis of the age patterns for ulcerative colitis shows certain uniform characteristics in seven clinical series from England (three), New Zealand, Norway, and the United States (two).

The disease appears to be confined to genetically predisposed persons. Two groups can be distinguished. Early-onset patients differ genetically and clinically from late-onset patients. Regularities in the world-wide epidemiological data can be explained if the disease is initiated in predisposed people by intrinsic random events. We propose, following Burnet, that these events correspond to a specific form of somatic gene mutation in mesenchymal stem cells. These spontaneous gene changes initiate the growth of a 'forbidden clone' of cells whose 'mutant' humoral products attack the colonic mucosa. Various extrinsic factors can influence the growth and activity of the forbidden clone by affecting an endogenous defence system. Exacerbations and remissions in the clinical course of the disease depend on the waxing and waning of the forbidden clone.

This general class of disease has been described as 'disturbed-tolerance autoimmune'. However, the serological evidence indicates that the primary pathogenic agents, or 'autoantibodies', are unlikely to be immunoglobulins. We propose that 'autoaggressive' is a more appropriate term than 'autoimmune'.

REFERENCES


Basophilic granulocytes in inflammatory tissues of man

Rupe, C. E. (1963). Mast cells in non-neoplastic diseases of man and
their relationship to the basophilic leukocyte. Ibid., 103, 436-440.

factor autoantibody and intrinsic factor in man by radio-
immunodiffusion and radioimmunoelectrophoresis. J.
Immunol., 95, 536-541.

of patients with chronic ulcerative colitis: a review of 2000.
cases. Gastroenterology, 16, 25-38.

med., 26, 542-553.

Ustvedt, H. J. (1958). Ulcerative colitis in Norway, in recent studies in
epidemiology (ed. J. Pemberton and H. Willard), Blackwell,
Oxford.

J., 1, 1447.

colitis in New Zealand, showing a low incidence in Maoris.
Brit. med. J., 2, 228-231.

Wilcox, R. G., and Isselbacher, K. J. (1961). Chronic liver disease in
young people. Clinical features and course in thirty-three

Wright, R., Lumsden, K., Luntz, M. H., Sevel, D., and Truelove,
S. C. (1965). Abnormalities of the sacro-iliac joints and

Wright, V., and Watkinson, G. (1965a). The arthritis of ulcerative

——, —— (1965b). Sacro-iliitis and ulcerative colitis. Ibid., 2,
675-680.
Aetiology of ulcerative colitis. II. A new hypothesis.

P R Burch, F T de Dombal and G Watkinson

Gut 1969 10: 277-284
doi: 10.1136/gut.10.4.277

Updated information and services can be found at:
http://gut.bmj.com/content/10/4/277.citation

Email alerting service

These include:
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections
Articles on similar topics can be found in the following collections
Ulcerative colitis (1113)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/