

Estimating the incidence of coeliac disease with capture-recapture methods within four geographic areas in Italy

G Corrao, P Usai, G Galatola, N Ansaldi, A Meini, M A Pelli, G Castellucci, G R Corazza and the Working Group of the Italian "Club del Tenue"

Abstract

Study objective – To estimate the incidence rate of newly diagnosed cases of coeliac disease in Italy.

Design – This was a descriptive study of coeliac disease incidence in the period 1990–91.

Setting – During 1990–91 newly diagnosed cases of coeliac disease were signalled by several sources including diagnostic records of departments of paediatrics, general medicine and gastroenterology, national health service records for the supply of gluten free diets and the archives of the Italian Coeliac Society.

Patients – Altogether 1475 cases were flagged throughout Italy, 478 of whom were selected, corresponding to 270 individual patients from a target population resident in four areas: Provinces of Turin and Cuneo (Piedmont Region, northern Italy); Province of Brescia (Lombardia Region, northern Italy); Umbria Region (central Italy) and Sardinia Region (insular Italy). Only for these areas were patients flagged from several sources and the reference population was identifiable.

Main results – The overall crude incidence rates for all ages per 100 000 residents per year were 2.4, 2.7, 1.5, and 1.7 in the four areas, respectively. The childhood cumulative incidence rates (aged ≤ 15 years) per 100 000 live births were 143, 141, 72, and 80 respectively. The mean ages at diagnosis were similar for both childhood and adult cases throughout the areas – these were around 4 and 34 years respectively. For each area, the incidence rate was constantly higher in the main city than elsewhere. Using the capture-recapture method, an estimated completeness of case archives of 0.84 was obtained, whereas this figure was only 0.47 for hospital sources.

Conclusions – This population based study on the incidence of coeliac disease shows that several information sources should be used to avoid underestimation. The incidence rate of coeliac disease in Italy was among the highest in Europe, and was widely variable showing highest figures in Piedmont and Lombardia and the lowest in Umbria and Sardinia. This trend was not due to different age at diagnosis, which suggests variable diagnostic awareness

of the disease rather than different environmental patterns affecting the clinical presentation.

(*J Epidemiol Community Health* 1996;50:299–305)

Although coeliac disease (CD) is a health problem that carries an increased risk of malignancy, epidemiological data concerning its incidence are still incomplete. High frequencies of clinically manifest disease in children have been reported from western Ireland¹ and Sweden,² whereas the lowest reported childhood incidence rates come from Finland³ and Denmark.⁴ No information is available from the United States of America⁵ and few reports are available for Mediterranean countries.^{6–8} This epidemiological information is valuable as it helps to improve understanding of aetiological factors and the implementation of health programmes aiming towards better recognition and treatment of the disease.⁹

In the past two decades the timing of diagnosis has gradually moved towards adulthood¹⁰ because of an increasing recognition of atypical and subclinical cases.¹¹ Nonetheless, the incidence of CD has been assessed, with a few exceptions,^{12–14} only in children.

Epidemiological evidence of different geographical and/or temporal variability in the incidence of CD is difficult to interpret. This is because of wide variability in clinical awareness,¹⁵ endoscopic duodenal biopsy,¹⁶ and implementation of serological screening.¹⁷ Moreover, in the absence of a national register of the disease, several information sources should be used to measure incidence, since the use of single sources such as death certificates, CD societies, or hospital records have been shown to yield a quarter of the actual cases.¹⁸

These considerations prompted us to constitute a working group within the Italian Association for the Study of Small Bowel Diseases ("Club del Tenue"), with the aim of assessing the incidence of newly diagnosed cases of coeliac disease in Italy. This paper reports the results of the study for four well defined geographic areas.

Methods

CATCHMENT AREAS

The overall number of CD cases flagged by all participating centres was 1475. We selected only those areas where flagging satisfied the criteria indicated below in "data collection".

Institute of Statistical and Mathematical Sciences, University of Milan, Italy
G Corrao

Department of Medicine, University of Cagliari, Italy
P Usai

Division of Gastroenterology, Ospedale Mauriziano Umberto I, Turin, Italy
G Galatola

Department of Paediatrics, University of Turin, Italy
N Ansaldi

Department of Paediatrics, University of Brescia, Italy
A Meini

Department of Gastroenterology, University of Perugia, Italy
M A Pelli

Department of Paediatrics, University of Perugia, Italy
G Castellucci

Department of Internal Medicine, University of L'Aquila, Italy
G R Corazza

Correspondence to:
Prof G Corrao,
Chair of Medical Statistics and Epidemiology,
Institute of Statistical and Mathematical Sciences "Marcello Boldrini",
University of Milan,
Via Conservatorio 7
Milano, Italy

Accepted for publication
October 1995

Thus, the study includes cases of CD newly diagnosed between 1 January 1990 and 31 December 1991 in the resident population of four Italian areas: the provinces of Turin and Cuneo (Piedmont Region), the Province of Brescia (Lombardia Region), the Umbria Region, and the Sardinia Region. In these areas, according to the 1991 Italian population census the resident population was 6 339 194 (11.0% of the whole Italian population) and 107 048 live births were recorded during the study period (9.4% of all live births in Italy).

DATA COLLECTION

CD patients were identified from four different information sources:

- Diagnostic lists of paediatric, general medicine and gastroenterology departments of hospitals in the study areas;
- Diagnostic lists of leading Italian hospitals likely to attract CD patients throughout the country;
- National health service records of patients for whom gluten free food was provided on the basis of histological evidence of CD;
- Archives of the local branches of the Italian Coeliac Society.

Data of patients entered for the first time during the study period were independently collected from each of the four sources considered. Details included surname, name, date of birth, date of first diagnosis, birthplace, and residence area. Flagging archives were thus obtained where individual patients were expected to be reported more than once, indicating flagging from multiple sources. Record linkage by surname, name, gender, and date of birth allowed us to build archives of newly diagnosed cases of CD during the study. We excluded from the study patients whose name reappeared in the records after interruption of their gluten free diet and patients who were already present in the national health service or Italian Coeliac Society records in the years preceding the study period (prevalent cases), in whom the diagnosis had therefore already been made. In order to calculate the incidence of new cases within the selected geographical areas, patients diagnosed within these areas but resident elsewhere were also excluded.

EVALUATING DIAGNOSTIC APPROACH

The diagnosis of CD for all patients included in the study was performed by retrospective examination of data collected routinely in the hospitals where the diagnoses were made. Patients were classified into three categories according to the diagnostic approach as follows: diagnosis unsupported by duodenal or jejunal biopsy; based on a single abnormal duodeno-jejunal histology finding; on diagnosis formulated after the European Society of Paediatric Gastroenterology and Nutrition (ESPGAN) criteria. The latter entails abnormal duodenal or jejunal histology followed by clinical or histological return to normal on gluten free diet.¹⁹

ESTIMATING COMPLETENESS OF CASE ARCHIVES AND ITS SOURCES OF VARIABILITY

To estimate completeness of case ascertainment for the whole archives and each information source we used the Lincoln-Peterson capture-recapture method,^{20,21} which compares results for multiple independent sources ascertaining the same event. Completeness of the case archives was expressed as the proportion between the observed and expected numbers of new diagnosis in the target population (N). To estimate N, we considered hospital flagging (sources 1 and 2 of data collection section) as primary sources, while national health service and Italian Coeliac Society were considered as secondary sources. N was calculated according to the Chapman estimator²²:

$$N = \frac{(M+1)(n+1)}{(m+1)} - 1 \quad (1)$$

where M is the number of new diagnoses identified by the primary sources; n is the number of new diagnoses identified by at least one of the secondary sources; and m is the number of new diagnoses identified by both the primary and at least one of the secondary sources. An approximate unbiased estimate of the variance of N was derived by Seber²³ and given as:

$$\text{Var}(N) = \frac{(M+1)(n+1)(M-m)(n-m)}{(m+1)^2(m+2)} \quad (2)$$

The 95% confidence interval (CI) of N was calculated using the formula:

$$95\% \text{CI} = \pm 1.96 \sqrt{\text{Var}(N)} \quad (3)$$

Equation (1) produces a valid estimate of the expected number of cases under the assumption of independence of data sources, – that is, each subject must have an equal chance of being reported as case in the secondary sources, regardless of whether or not he/she was identified as case by the primary sources.

We then constructed a matrix that considers the number of new diagnoses reported in the archives and the estimated number of unreported new diagnoses for each possible combination of gender, age (≤ 15 v > 15 years), area of residence, and residence within or outside the area's main city. The estimated number of unreported diagnoses was calculated by the difference between expected cases (equation 1) and those reported in the archives. This structure corresponds to a case-control study, where cases and controls are the patients unreported and reported in the archives respectively. The risk of unreporting associated with the considered variables was estimated by a logistic regression model and expressed as the maximum likelihood estimate of relative risks and their 95% confidence intervals.²⁴ These calculations were performed using the logistic procedure of the SAS package.²⁵

Table 1 Completeness of reporting newly diagnosed patients in the four Italian areas

Completeness indexes	Piedmont	Lombardia	Umbria	Sardinia	Total
M*	78	34	12	27	151
m†	54	23	8	19	104
n‡	110	46	20	47	223
N§	158	68	29	66	322
(95%CI)¶	(142,175)	(57,78)	(21,37)	(54,78)	(298,348)
Cases archive completeness**	0.85	0.84	0.82	0.83	0.84
Hospital source completeness††	0.49	0.50	0.41	0.41	0.47

Piedmont = Turin and Cuneo Provinces; Lombardia = Brescia Province; Umbria & Sardinia = entire region; * number of new diagnoses identified in the diagnostic lists of hospital departments (primary source); † number of new diagnoses identified in both the primary and at least one of secondary sources; ‡ number of new diagnoses identified in at least one of the secondary sources; § number of new diagnoses estimated according to equation 1 ("Methods" section); ¶ 95% confidence interval of N (equation 3 in "Methods" section); ** proportion of observed number of new diagnoses over those expected (N); †† proportion of new diagnoses identified in the diagnostic list of hospital departments; (M) over the expected number of new diagnoses (N).

ESTIMATING OVERALL INCIDENCE RATE AND ITS SOURCES OF VARIABILITY

The overall crude incidence rate was calculated as the number of new diagnoses referred to the two years of observation divided by twice the 1991 census population, and expressed as the average number of cases per 10^{-5} residents per year. Incidence rates were calculated using both the observed and expected number of new diagnoses and the number of new diagnoses flagged only by hospital sources.

The relationship between the observed overall incidence rate and gender, age, and area of residence was assessed by a Poisson's multiple linear regression model.²⁶ This considers the incidence rate as the dependent variable, and gender, age, area of residence, and residence within or outside the area's main city as the independent variables. We then calculated the maximum likelihood estimate of the relative risk and the corresponding 95% confidence interval for each level of independent variables. These calculations were performed using the EGRET package.²⁷

ESTIMATING CHILDHOOD CUMULATIVE INCIDENCE

The childhood cumulative incidence was not directly computable since our observation period covered only two years. We therefore used the density method²⁸ based on the functional relationship existing between incidence rate and childhood cumulative incidence:

$$CCI_{0-j} = 1 - \exp(-\sum_j IR_{0-j}) \quad (4)$$

where CCI_{0-j} is the cumulative incidence occurring between birth and age j , and $\sum_j IR_{0-j}$ sums the age specific annual incidence rates occurring between the first and the j th year of age. Incidence rates used in equation (4) were obtained from the observed number of cases. The childhood cumulative incidence was computed for each of the four areas, and individually for each main city, between the 1st and the 15th year of age and expressed as the number of cases occurring from birth to the j th year of age in a hypothetical cohort of 100 000 live births. Equation (4) produces a valid estimate of the childhood cumulative incidence under the assumption that the incidence rate remains constant during the calendar period from the age of birth of the hypothetical cohort and the year of observation of cases.

EVALUATING SOURCES OF VARIABILITY OF AGE AT DIAGNOSIS

A hierarchical model of analysis of the variance (ANOVA) for unbalanced data²⁹ was fitted to evaluate the sources of variability of age at diagnosis. In the model, the main effects of gender and residence areas and the nested effect of city of residence (main city or other cities) within the residence area, were considered. The model was applied in the two strata corresponding to age ≤ 15 years and >15 years, arbitrarily considered as paediatric and adult ages respectively. These calculations were performed using the GLM procedure of the SAS package.³⁰

Results

FROM FLAGGING ARCHIVES TO CASE ARCHIVES

We received 478 flagged records of newly diagnosed CD patients during 1990–91 from the four areas included in the study. The hospital lists provided 181 records (151 from the study areas and 30 from other hospitals outside the areas); the national health service and the Italian Coeliac Society lists provided 220 and 77 records respectively. From the 478 flagged records we constructed a case archives of 270 patients, with an average per patient flagging frequency of 1.77.

Among the 270 patients considered in the study, 89 were males (male:female ratio 1:2.0). Altogether 139 patients were diagnosed at ≤ 15 years and 131 at >15 years. The overall mean age at diagnosis was 21 years; it was 3.7 for paediatric cases (≤ 15 years) and 34 years for adult cases (>15 years).

DIAGNOSTIC APPROACH

Of the 270 patients, 231 (85.6%) had medical records available. Their diagnosis was always supported by duodenal or jejunal biopsy: in 109 cases (47.2%) diagnosis was made according to a single abnormal duodeno-jejunal histology finding and in the remaining 122 cases according to ESPGAN criteria.

COMPLETENESS OF CASE ARCHIVES AND ITS SOURCES OF VARIABILITY

Data used in calculating expected cases in the four areas are reported in table 1. Of 151 cases flagged by hospitals (M), 104 were also flagged by at least another source (m) and 223 cases

Table 2 Relationship between risk of unreporting new diagnoses and gender, age and residence area

Independent variables	Categories	Unreported No (%)	Reported No (%)	RR*	(95% CI)†
Gender:	Males‡	18(34.4)	89(33.0)	1.00	—
	Females	33(65.6)	181(67.0)	0.90	(0.48,1.69)
Age classes (y):	≤15‡	28(55.7)	139(51.5)	1.00	—
	>15	23(44.3)	131(48.5)	0.87	(0.48,1.59)
Residence area:	Piedmont‡	24(47.1)	134(49.6)	1.00	—
	Lombardia	11(21.6)	57(21.1)	1.08	(0.49,2.35)
	Umbria	5(9.8)	24(8.9)	1.16	(0.40,3.35)
	Sardinia	11(21.6)	55(20.4)	1.12	(0.51,2.44)
City of residence:	Main city‡	17(32.8)	94(34.8)	1.00	—
	Other	34(67.2)	176(65.2)	1.07	(0.57,2.02)

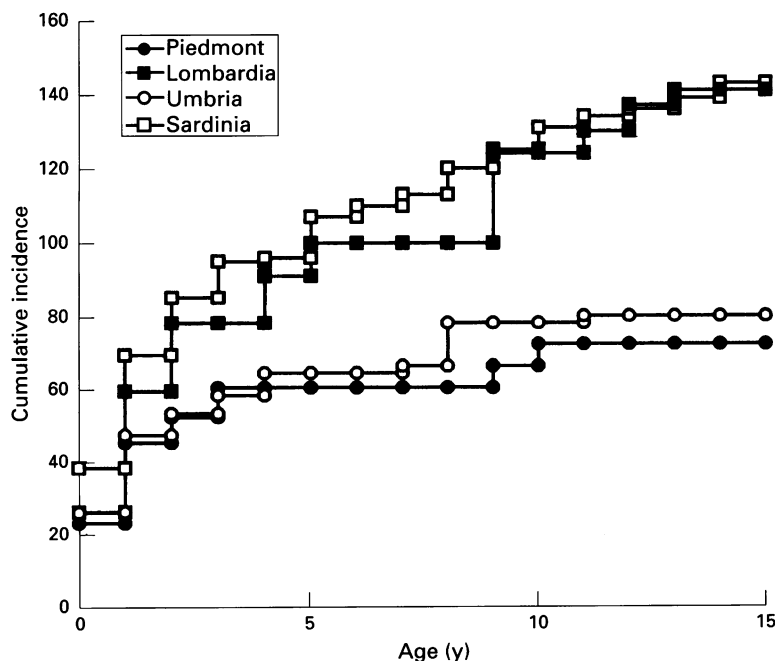
Piedmont = Turin and Cuneo Provinces; Lombardia = Brescia Province; Umbria and Sardinia = entire region; * Maximum likelihood estimates of the relative risk obtained by a logistic regression model; RR is the risk of unreporting new diagnoses within a category with respect to the reference category; † 95% confidence interval of RR; ‡ reference category.

Table 3 Crude incidence rates in the four Italian areas calculated either from hospital flagging alone, multiple sources information (observed cases) or estimated according to the capture-recapture methods (expected cases)

Residence area	Resident population*	Hospital flagging		Observed cases		Expected cases	
		No	IR†	No	IR†	No	IR†
Piedmont	2822	94	1.67	134	2.37	158	2.80
Lombardia	1040	41	1.97	57	2.74	68	3.27
Umbria	820	14	0.85	24	1.46	29	1.77
Sardinia	1657	32	0.97	55	1.66	66	2.00
Total	6339	181	1.43	270	2.13	322	2.54

Piedmont = Turin and Cuneo Provinces; Lombardia = Brescia Province; Umbria and Sardinia = entire region; * from 1991 Italian population census (per 10³); † incidence rates (per 10⁻³ per year⁻¹).

were flagged by at least one of the secondary sources (national health service and/or Italian Coeliac Society). Using these frequencies in equation 1, 52 cases (322–270) escaped from the case archives because of incomplete source information. The underestimation was homogeneous throughout the areas, ranging from 0.15 for Piedmont to 0.18 for Umbria. The completeness of hospital source ranged from



Childhood cumulative incidence (per 100 000 live births) in the four Italian areas.

Table 4 Relationship between observed incidence rate and gender, age, and residence area

Independent variables	Categories	RR†	(95% CI)‡
Gender:	Males*	1.00	—
	Females	1.90§	(1.48,2.45)
Age classes (y):	0–15*	1.00	—
	16–39	0.33§	(0.25,0.44)
	40–59	0.21§	(0.15,0.30)
	≥60	0.11	(0.06,0.18)
Residence area:	Sardinia*	1.00	—
	Umbria	1.24	(0.65,1.70)
	Lombardia	2.69§	(1.11,2.37)
	Piedmont	1.76§	(1.22,2.29)
City residence:	Main city	1.55§	(1.09,4.83)
	Other*	1.00	—

Sardinia and Umbria = entire region; Lombardia = Brescia Province; Piedmont = Turin and Cuneo Provinces; * Reference category; † maximum likelihood estimates of the relative risk obtained by a Poisson's multiple linear regression model; ‡ 95% confidence interval of RR; § p<0.05.

0.41 (Sardinia and Umbria) to 0.50 (Lombardia).

Table 2 shows that the estimated risk of unreported new diagnosis was not associated with gender, age, or area and city of residence.

OVERALL INCIDENCE RATE AND ITS SOURCES OF VARIABILITY

Table 3 shows that the crude incidence rate was highest in the northern areas (Piedmont and Lombardia) and lowest in the other areas, irrespective of whether it was calculated from hospital sources alone or from observed or expected cases.

Table 4 shows the estimated relative risks associated with gender, age, and area of residence. Females had a significantly higher relative risk. Despite a progressive reduction in the relative risk with increasing age, an appreciable number of cases were diagnosed above age 60 (n = 15). By taking cases resident in Sardinia as the reference category, relative risks were significantly higher in northern Italy. Cases resident in the main cities also showed a significantly higher relative risk.

CHILDHOOD CUMULATIVE INCIDENCE

The figure shows the observed childhood cumulative incidence values according to age in the four areas. Two clusters were present. For northern areas, the childhood cumulative incidence was constantly the highest, reaching twice the value of the other two areas at the age of 15 years.

Table 5 shows the estimated childhood cumulative incidence at the age of 2, 5, 10, and 15 years in the four areas and in the respective main cities. For all ages the values were highest in the northern areas and within each area the highest figures were observed for the respective main cities.

SOURCES OF VARIABILITY OF AGE AT DIAGNOSIS

Table 6 shows the results of the ANOVA performed separately for the two age groups. In both paediatric (≤15 years) and adult (>15

Table 5 Childhood cumulative incidence (per 100 000 live births) in the four Italian areas and in their respective main cities

Residence area		Age (y)			
		2	5	10	15
Piedmont	Entire area	69.3	96.6	125.4	143.2
	Turin*	87.0	123.6	169.3	180.7
Lombardia	Entire area	59.0	91.2	124.3	141.4
	Brescia*	83.6	125.7	125.7	194.3
Umbria	Entire area	44.7	60.1	66.4	72.4
	Perugia*	69.9	93.7	108.7	108.7
Sardinia	Entire area	47.0	63.7	77.8	79.8
	Cagliari*	75.3	102.8	120.7	120.7
Total	Entire areas	57.5	80.8	103.0	114.2
	Main cities	81.5	112.1	145.2	157.9

Piedmont = Turin and Cuneo Provinces; Lombardia = Brescia Province; Sardinia and Umbria = entire region. * Main cities.

Table 6 Sources of variability of age at diagnosis

Sources of variability	df	F value	p>F	Modalities	LS mean§	SE¶	
Age ≤15 y							
Gender*	1	1.53	0.218	Males	3.2	0.4	
				Females	3.9	0.3	
Area*	3	1.15	0.331	Piedmont	3.9	0.5	
				Lombardia	3.7	0.6	
				Umbria	3.7	0.7	
				Sardinia	3.4	0.6	
City†	4	1.08	0.369	Piedmont	Turin‡	4.0	0.8
					Other	3.8	0.5
				Lombardia	Brescia‡	3.5	0.9
					Other	3.8	0.4
				Umbria	Perugia‡	3.8	0.9
					Other	3.6	0.5
				Sardinia	Cagliari‡	3.4	0.9
					Other	3.4	0.5
Age >15 y							
Gender*	1	9.14	0.003	Males	30.4	4.1	
				Females	36.2	4.3	
Area*	3	1.36	0.260	Piedmont	35.3	4.5	
				Lombardia	32.8	4.4	
				Umbria	34.0	4.9	
				Sardinia	32.0	4.4	
City†	4	0.01	0.926	Piedmont	Turin‡	35.0	4.6
					Other	35.4	4.2
				Lombardia	Brescia‡	32.0	4.9
					Other	33.1	4.2
				Umbria	Perugia‡	33.7	5.1
					Other	34.6	4.8
				Sardinia	Cagliari‡	32.0	4.6
					Other	32.1	4.2

Piedmont = Turin and Cuneo Provinces; Lombardia = Brescia Province; Umbria and Sardinia = entire region. * Estimate of the main effect of the variable; † estimate of the nested effect of city of residence within the residence area; ‡ main cities; § least-square means; ¶ standard error of LS means.

years) patients, the age at diagnosis was significantly homogeneous according to the area of residence and to the city of residence within each area. A significantly higher age at diagnosis was observed in women.

Discussion

Epidemiological studies of CD pose several problems, particularly when they are trying to estimate its incidence rate.³¹ Firstly, all diagnoses originating from well defined populations should be included to estimate the disease incidence correctly. The total flagged records in this study were 1475, but only 478 originated from well defined geographical areas. For most participating centres, flagging was obtained only from some of the hospital departments in the areas; several health districts were not able to provide the requested information and most Italian Coeliac Society branches had no authorisation to provide confidential patient data. A well defined originating population was found only for four areas, covering 10% of both

the Italian population and the number of live births in the study period.

We may have introduced bias by selecting areas in which the greatest diagnostic awareness existed. However, this should not have affected the validity of comparing results between these areas. Incomplete inclusion of all diagnoses might occur, as shown by the finding that 52% of our patients were diagnosed outside their residence area. To avoid this, we used multiple information sources which also included referral hospitals outside the study areas likely to attract patients throughout the country. These provided only 6.3% of all flagged records, however, whereas 62% of them were derived from specific national health service lists and branches of the Italian Coeliac Society and would have been missed using only hospital sources. In areas where hospital flagging was scarce, we observed a balance due to high flagging rates from the other sources. The use of the capture-recapture technique²⁰⁻²³ estimated that our approach recruited 84% of CD patients newly diagnosed in defined reference populations. This 16% underestimate of the incidence rate was uniformly distributed among the areas and was independent of gender, age and rural or urban residence. Our comparisons are therefore correct and unbiased by under-reporting according to the type of populations. Our calculated childhood cumulative incidence in Piedmont was twice the figure reported for that area by the paediatric multicentre ESPGAN study ($1.4/10^{-3}$ v $0.7/10^{-3}$ respectively),⁸ which used only hospital sources. If we consider data from our hospital source, we actually obtain the same figure as that reported by the ESPGAN study. Thus, we believe that our data are more likely to reflect the actual coeliac disease incidence.

The second problem is that standard diagnostic criteria should be used. We verified that for all our patients whose medical records could be retrieved, diagnoses were formulated according to histological criteria, although only in half the cases were the strictest ESPGAN criteria used.¹⁹ This should not have biased our results, since diagnosis of CD formulated according to a single histological finding, is confirmed according to the ESPGAN criteria in 95% of the cases.³²

Thirdly, diagnostic criteria should be homogeneous throughout the country and the period considered. We observed no difference in diagnostic criteria between the four areas and according to urban or rural residence within each area. Moreover, our study covered an observation period of just two years, to avoid bias due to time-dependent changes in the diagnostic approach. Thus, we were not able to measure directly the childhood cumulative incidence, and based our estimates on the density method that exponentially transforms age specific rates of a dynamic population observed transversally. The validity of our estimates depends on assuming a constant age specific incidence rate throughout the years. In the last 20 years, the frequency of diagnosing sub-clinical CD in Italy has increased,¹⁵ suggesting an underestimate of our figures. However, this

increased diagnostic frequency in a long lasting observational study may be affected by changes in diagnostic modalities and/or awareness rather than reflecting an actual increased disease incidence.

Fourthly, in an incidence study cases are included when the diagnosis is formulated, and the onset of disease cannot be dated. The calculation of the incidence rate is therefore of dubious validity,³¹ particularly for CD whose onset is often gradual and escapes dating. Methods of measuring cumulative incidence are implemented to estimate the actual frequency of disease occurrence. A declining childhood cumulative incidence trend during the 1980s has been reported from Ireland and United Kingdom^{33,34} and other European areas.³⁵ Increased mean age at diagnosis may explain this finding,¹⁰ pointing to a need to include diagnoses formulated at all ages and not only during childhood. Possible temporal trends and/or between-population variability of diagnostic frequency will therefore be appreciated. Cautious interpretation of incidence rates referring to adulthood is needed, as these reflect the diagnostic frequency of the disease rather than its actual incidence, particularly for our study where it was not possible to assess symptoms leading to the diagnosis and their time of onset.

Lastly, in incidence studies only newly diagnosed cases are included and these do not necessarily coincide with newly occurring cases. Due to the wide clinical spectrum of CD, the observed variability in disease frequency between the areas may indicate variable rate of detection rather than variable incidence rate.

Our results are comparable with the highest disease frequency reported in northern European countries.³⁶ Our estimated childhood cumulative incidence was the second highest for Europe after Sweden,² and showed a wide geographical variation between the lowest figure of $0.7/10^{-3}$ in Umbria (central Italy) and the highest of $1.4/10^{-3}$ in Piedmont (northern Italy). Consistently higher figures were also found for each area in children living in the main cities compared with those living elsewhere. A role for environmental factors, such as (1) low rate of breast feeding,^{37,38} (2) early dietary introduction of gluten proteins and high amount of gluten in the weaning diet,³⁹⁻⁴¹ (3) low incidence of gastroenteritis³⁴ may be suggested. These factors should, however, determine an earlier onset of disease symptoms, whereas our cases had a similar mean age at diagnosis in the four areas studied. The highly variable childhood cumulative incidence observed may possibly be due to increased diagnostic awareness and availability of diagnostic facilities in the largest cities in the northern areas.

Our adult incidence rate followed the same distribution as the childhood cumulative incidence. This further supports the hypothesis of a different diagnostic awareness of the disease, since it is unlikely that environmental factors are kept constantly different throughout the areas for all birth cohorts of the patients included in our study.

In conclusion, we have shown that population based studies of the incidence rate of CD should use several information sources to avoid underestimation. By this approach, we showed that the childhood incidence rate in Italy is among the highest in Europe, and that wide variability is observed between geographical areas for both the paediatric and adulthood incidence rates. The highest rates were also observed in metropolitan areas within each area. The age at diagnosis was homogeneous throughout the areas, suggesting that geographical variability probably depends more on different disease awareness than on varying environmental factors affecting the clinical presentation of CD.

The working group of the Italian "Club del Tenue" comprises: C Catassi (Ancona); C Panella (Bari); G Brusco (Bologna); M V Cherchi (Cagliari); G Bottaro (Catania); F Bascietto (Chieti); A Spada (Cuneo); G Gemme (Genova); P Torchio, A Pienabarca (L'Aquila); C Romano, G Magazzù (Messina); M T Bardella (Milano); S Guandalini, R Sollazzo (Napoli); F Cataldo, G Iacono (Palermo); C Leonardi (Roma); C Satagna Guidetti, D Dell'Olio, E Malorgio, G Ostino (Torino); G Mastella (Verona)

- 1 Mylotte M, Egan-Mitchell B, McCarthy CF, McNicholl B. Incidence of coeliac disease in the west of Ireland. *BMJ* 1973;i:703-5.
- 2 Ascher H, Krantz I, Kriatianson B. Increasing incidence of coeliac disease in Sweden. *Arch Dis Child* 1991;66:608-11.
- 3 Mäki M, Holm K. Incidence and prevalence of coeliac disease in Tampere. *Acta Paediatr Scand* 1990;79:980-2.
- 4 Weile B, Krasilnikoff PA. Low incidence rates by birth of symptomatic coeliac disease in Danish population of children. *Acta Paediatr Scand* 1992;81:394-8.
- 5 Trier JS. Celiac sprue. *N Engl J Med* 1991;325:1709-19.
- 6 Mouterde O, Chafai S, Amar R, Leluyer B, Mallet E. Profil épidémiologique de la maladie coéliquie en Seine-Maritime. *Arch Fr Pédiatr* 1993;50:539-40.
- 7 Dahan S, Slater PE, Cooper M. Coeliac disease in the Rehovot-Ashdod region of Israel: incidence and ethnic distribution. *J Epidemiol Community Health* 1984;38:58-60.
- 8 Greco L, Di Donato F, Ansaldo N, et al. Analysis of incidence trends: an example from the Italiana cohort. In: Auricchio S, Visakorpi JK, eds. *Common food intolerances. 1: Epidemiology of coeliac disease*. Basel: Karger, 1992;45-56.
- 9 Langman MJS. Can epidemiology help us to prevent celiac disease? *Gastroenterology* 1986;90:489-90.
- 10 Kelly DA, Phillips AD, Elliott EJ, Dias JA, Walker-Smith JA. Rise and fall of coeliac disease 1960-1985. *Arch Dis Child* 1989;64:1157-60.
- 11 Logan RFA, Tucker G, Rifkind EA, Heading RC, Ferguson A. Changes in clinical features of coeliac diseases in adults in Edinburgh and the Lothians 1960-1979. *BMJ* 1983;286:95-7.
- 12 Langman MJS, McConnell TH, Spiegelhalter DJ, McConnell RB. Changing patterns of coeliac disease frequency: an analysis of coeliac society membership records. *Gut* 1985;26:175-8.
- 13 Logan RFA, Rifkind EA, Busuttill A, Gilmour HM, Ferguson A. Prevalence and "incidence" of coeliac disease in Edinburgh and the Lothian region of Scotland. *Gastroenterology* 1986;90:334-42.
- 14 Jansen TLT, Mulder CJJ, Karssen PHZ, Wagenaar CGJ. Epidemiological survey of the Dutch Coeliac Disease Society: an update 1992. *Eur J Gastroenterol Hepatol* 1993;5:73-8.
- 15 Corazza GR, Frisoni M, Treggiari EA, et al. Subclinical celiac sprue. Increasing occurrence and clues to its diagnosis. *J Clin Gastroenterol* 1993;16:16-21.
- 16 Mee AS, Burke M, Newman J, Cotton PB. Small bowel biopsy for malabsorption: comparison of diagnostic adequacy of endoscopic forceps and capsule biopsy specimens. *BMJ* 1985;291:769-72.
- 17 Corrao G, Corazza GR, Andreani ML, et al. Serological screening of coeliac disease: choosing the optimal procedure according to various prevalence values. *Gut* 1994;35:771-5.
- 18 Loft DE. The epidemiology of diagnosis of coeliac disease. *Eur J Gastroenterol Hepatol* 1993;5:69-72.
- 19 Walker-Smith JA, Guandalini S, Schmitz J, Shmerling DH, Visakorpi JK. Revised criteria for diagnosis of coeliac disease. Report of working group of European Society of Paediatric Gastroenterology and Nutrition. *Arch Dis Child* 1990;65:909-11.
- 20 Cormack RM. The statistics of capture-recapture methods. *Oceanography and Marine Biology Annual Review* 1968;6:455-506.
- 21 Cochi SL, Edmonds LE, Dyer K, et al. Congenital rubella syndrome in the United States 1970-1985. *Am J Epidemiol* 1989;129:349-61.
- 22 Chapman DG. *Some proprieties of the hyper-geometric dis-*

- tribution with applications to zoological sample census. Berkeley: University of California Publications in Statistics, 1951;131-60.
- 23 Seber GAF. *The effect of trap response on tag recapture estimates.* *Biometrics* 1970;26:13-22.
 - 24 Breslow NE, Day NE. *Statistical methods in cancer research. Vol. I. The design and analysis of case-control studies.* IARC Scientific Publication No. 32. Lyon: International Agency for Research on Cancer, 1980.
 - 25 Harrell FE. The logistic procedure. In: *Sugi supplemental library users' guide.* Version 5 ed. Cary NY: SAS Institute Inc. 1986;269-93.
 - 26 Breslow NE, Day NE. *Statistical methods in cancer research. Vol. II. The design and analysis of cohort studies.* IARC Scientific Publications No. 82. Lyon: International Agency for Research on Cancer, 1987.
 - 27 EGRET reference manual. *Statistics and epidemiology Research Corporation and Cytel Software Corporation,* 1990.
 - 28 Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiological research. Principles and quantitative methods.* New York: Van Nostrand Reinhold Company, 1982.
 - 29 Snedecor GW, Cochran WG. *Statistical methods.* 8th Ed. Iowa: Iowa State University Press, 1989.
 - 30 Statistical Analysis System. *SAS user's guide: statistics.* Version 5 ed. Cary, NY: SAS Institute Inc, 1985.
 - 31 Logan RFA. Problems and pitfalls in epidemiology studies of coeliac disease. In: Auricchio S, Visakorpi JK, eds. *Common food intolerances. 1: Epidemiology of coeliac disease.* Basel: Karger, 1992;14-24.
 - 32 Guandalini S, Ventura A, Ansaldi N, et al. Diagnosis of coeliac disease: time for a change? *Arch Dis Child* 1989; 84:1320-5.
 - 33 Littlewood JM, Crollick AJ. Childhood coeliac incidence is disappearing. *Lancet* 1980;ii:1359.
 - 34 Stevens FM, Egan-Mitchell B, Cryan E, McCarthy CF, McNicholl B. Decreasing incidence of coeliac disease. *Arch Dis Child* 1987;62:465-8.
 - 35 Greco L, Mäki M, Di Donato F, Visakorpi JK. Epidemiology of coeliac disease in Europe and Mediterranean area. A summary report on the multicentre study by the European Society of Paediatric Gastroenterology and Nutrition. In: Auricchio S, Visakorpi JK, eds. *Common food intolerances. 1: Epidemiology of coeliac disease.* Basel: Karger, 1992;25-44.
 - 36 Greco L, Mäki M, Di Donato F, Visakorpi JK. Epidemiology of coeliac disease in Europe and Mediterranean area. A summary report on the multicentre study by the European Society of Paediatric Gastroenterology and Nutrition. In: Auricchio S, Visakorpi JK, eds. *Common food intolerances. 1: Epidemiology of coeliac disease.* Basel: Karger, 1992;25-44.
 - 37 Auricchio S, Follo D, de Ritis G, Marubini E. Does breast feeding protect against the development of clinical symptoms of coeliac disease in children. *J Pediatr Gastroenterol Nutr* 1983;2:428-33.
 - 38 Greco L, Mayer M, Grimaldi M, Follo D, De Ritis G, Auricchio S. The effect of early feeding on the onset of symptoms in coeliac disease. *J Pediatr Gastroenterol Nutr* 1985;4:52-5.
 - 39 Kristiansson B, Ascher H, Odenman I, Sandberg L. The incidence of coeliac disease and changes in gluten consumption. In: Mearin ML, Mulder CJJ, eds. *Coeliac disease.* Amsterdam: Kluwer Academic Publishers, 1991;107-4.
 - 40 Mäki M, Holm K, Ascher H, Greco L. Factors affecting clinical presentation of coeliac disease: role of type and amount of gluten-containing cereals in the diet. In: Auricchio S, Visakorpi JK, eds. *Common food intolerances. 1: Epidemiology of coeliac disease.* Basel: Karger, 1992;76-82.
 - 41 Michaelsen KF, Weile B, Larsen P, Samuelson G, Krasilnikoff PA. Does the low intake of wheat in Danish infants cause the low incidence rate of coeliac disease? *Acta Paediatr Scand* 1993;82:605-6.