Incomplete overlapping of biological, clinical, and environmental information in molecular epidemiological studies: a variety of causes and a cascade of consequences

M Porta, N Malats, J Vioque, A Carrato, M Soler, L Ruiz, V Barberà, D Ayude, F X Real

J Epidemiol Community Health 2002; 56:734–738

Recent scientific, political, industrial, and public relations efforts to publicise the first interpretations of the human genome sequence raised unprecedented social expectations towards the health sciences; yet, prudence seems warranted when assessing the relevance of emerging genomic knowledge to lessen the burden of illness in human populations. Among the formidable tasks that lie ahead, it will be important to achieve a proper understanding of how genes and the environment interact to cause human disease. This aim will partly depend on how molecular epidemiological studies are conducted. Long viewed as a promising approach, molecular epidemiology is no longer in its infancy. Figure 1 shows that the number of studies ascribed to molecular epidemiology increased dramatically during the past decade. Yet, as shown in table 1 and figure 2, “molecular epidemiological” studies are just a fraction of studies attempting to integrate epidemiological, genetic, and molecular dimensions of disease. Even if nowadays some features of molecular epidemiological studies retain some novelty, other basic challenges have always been undisputed. Prominently, the need to integrate valid biological, clinical, and epidemiological information from unbiasedly selected populations. Indeed, tumour tissue procurement and, more generally, the availability of biological samples are crucial in molecular epidemiology. However, although completeness of the “biological study base” is frequently hampered by ethical, clinical, and logistic factors, these are rarely analysed in depth. Unaccountably, molecular epidemiological studies often fail to properly assess whether selection biases affect their results. Exceptions do exist, of course. One is the study by Slattery et al—interestingly, published in Mutation Research Genomics. The authors were able to collect blocks and extract DNA for 2117 of 2477 people (86%) for whom they had permission to obtain tumour blocks. They found no differences in age, tumour site, or diet and lifestyle characteristics between subjects with and without DNA extracted. However, they were less likely to be able to extract DNA if the case was diagnosed at a more advanced disease stage or at the earliest stage.

INCOMPLETE OVERLAPPING OF BIOLOGICAL, CLINICAL, AND ENVIRONMENTAL INFORMATION: AN EXAMPLE

Studies providing opportunities to think about the consequences of an incomplete overlapping of biological, clinical, and environmental information abound. Of course, they include our own, where we have often tried to control the problem—not always successfully—and to discuss such consequences. Another recent example is the case-case study by Slebos et al. It investigated whether in pancreatic cancer alterations in the K-ras and p53 genes may be related

Figure 1 Number of articles from “molecular epidemiological” studies [column A of the table].
with aspects of the patients’ medical history, the histopathology of tumours, and environmental exposures. The study deserves to be commended on a number of strengths, including the use of advanced techniques for the molecular analyses of pure populations of tumour cells, the attempt to account for subjects that could not be included in genetic and organochlorine compounds analyses, and the cautious conceptualisation of the results. Such strengths notwithstanding, the study has potentially important limitations. The causes and consequences of such weaknesses merit analysis for two main reasons: firstly, because some of the study conclusions might be erroneous; and secondly, of even broader interest, because the limitations may illustrate some potential pitfalls of molecular epidemiological research.

This debate concerns many scholars who work on molecular epidemiology, and very directly those who study pancreatic cancer and other cancers whose diagnosis is especially difficult. To broadly contextualise the debate, the following facts are worth noting: over 12 years after their high prevalence at diagnosis was first reported, point mutations in the K-ras gene remain the most consistent genetic alteration in exocrine pancreatic cancer. Even more important, K-ras mutations are one of the most common genetic alterations in human cancers—probably, the most frequent somatic (acquired) oncogene mutation. It is hence puzzling that so little evidence is available for or against the hypothesis that the occurrence and persistence of K-ras mutations may be related to environmental factors. Let us then synthesise the main limitations of the study by Slebos et al. and analyse their potential effects upon their conclusions.

- Out of some 611, “more than 550”, or 374 (321 + 53) eligible patients, 61 (16%), 49 (13%), and 24 (6%) were integrated in the analyses of K-ras, p53 and organochlorine compounds, respectively (percentages are for 374). Because the percentage of patients potentially eligible for inclusion who were finally analysed is low by conventional epidemiological standards, it seems pertinent to ask why. A possible explanation regards the fact that about 53% or more of cases were deceased when contacted. This possibility should make us think about methods we use to identify eligible patients, and about practical ways to shorten the distance between the clinicians who diagnose cases and the epidemiologists-researchers. It is also possible that the low percentage of cases analysed was partly attributable to the fact that about two thirds of patients without surgical tumour material were excluded. If so, maybe the authors should have kept in mind the possibility of retrieving tumour material were excluded.

- Slebos et al. provided an interesting comparison of some characteristics of patients included (as their tumour blocks were available, n=61) and of patients included only in the original case-control study based on cases selected from the San Francisco Bay Area (n=321). It is unclear, however, how many of the 61 cases originated from the 321 cases, as some of the 61 patients came from 53 additional cases later retrieved from the University of California at San Francisco (UCSF) pathology records. It is also unclear if the latter cases were interviewed in a similar way as the former. It would also have helped (to assess the validity of results on organochlorine compounds) that the authors had provided data on their respective times of blood sampling (which apparently occurred about three months after diagnosis in the original study). These issues concern a large number of molecular epidemiological studies that tend to use patients from a variety of clinical settings in order to palliate the low number of cases with biological samples available to researchers.

Figure 2 Number of articles with the terms mentioned in columns A, B, and C of the table.
It would have been useful for Slebos et al to compare several other characteristics of cases included and excluded; prominently, tumour stage at diagnosis (or at the time of blood extraction, if available). Staging information is essential in any molecular epidemiological study of clinically aggressive diseases as pancreatic cancer. It is even more important when studying organochlorine compounds because lipid mobilisation (which might spuriously increase serum organochlorine concentrations) is more likely with increasing cancer dissemination.

We are not aware of scientific evidence supporting the claim that lipid levels may “stabilise” after a diagnosis of pancreatic cancer. We suggest it might also be of interest to have data on treatments given before blood sampling (because some might change serum organochlorine concentrations), changes in BMI (or weight loss), symptoms before blood collection, and lipid concentrations for the subsets of patients included and excluded from the original and the new study.

The comparison of characteristics could have been extended to the 49 and the 24 cases available, respectively, for p53 and organochlorine compounds analyses, as main conclusions of the study are based on the latter two groups. Briefly, the point is: while the attempts to describe subjects that could not be included in genetic and environmental analyses made by Slebos et al, Slattery et al, and ourselves are praiseworthy, the descriptions are often incomplete. The fact that most other studies do worse is of little comfort.

Can the previous problems on selection and information bias be of any consequence? The next set of issues suggests that they can.

In pancreatic cancer, tumour stage at diagnosis has consistently been found to be unrelated to K-ras mutation status by the larger studies. It is hence worth thinking why in the study by Slebos et al the 15 K-ras wild-type patients had more localised tumours than the 46 patients with K-ras mutations: the former were over three times more likely to have stage I tumours than the latter (OR = 3.86, 95% CI: 1.02 to 14.62). We see no clear explanation for this association. Perhaps a selection bias occurred, probably because of the small number of subjects. As said, the issue is also interesting because it may illustrate the cascade of consequences that selection bias may have for causal estimation. In this instance, the consequences may have been threefold.

The authors found that diabetes was more common among patients with wild type K-ras (six cases) than among patients with K-ras mutations (two cases). The finding was never reported before. But pancreatic cancer

<table>
<thead>
<tr>
<th>Year</th>
<th>&quot;Epidemiology, molecular&quot; A</th>
<th>&quot;Epidemiology&quot; and &quot;molecular&quot; B</th>
<th>&quot;Epidemiology&quot; and &quot;gene*&quot; C</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>216</td>
<td>1638</td>
<td>2912</td>
</tr>
<tr>
<td>1999</td>
<td>178</td>
<td>1343</td>
<td>2419</td>
</tr>
<tr>
<td>1998</td>
<td>204</td>
<td>1165</td>
<td>2108</td>
</tr>
<tr>
<td>1997</td>
<td>247</td>
<td>1175</td>
<td>2093</td>
</tr>
<tr>
<td>1996</td>
<td>210</td>
<td>1194</td>
<td>1817</td>
</tr>
<tr>
<td>1995</td>
<td>147</td>
<td>1173</td>
<td>1675</td>
</tr>
<tr>
<td>1994</td>
<td>73</td>
<td>963</td>
<td>1439</td>
</tr>
<tr>
<td>1993</td>
<td>49</td>
<td>778</td>
<td>1125</td>
</tr>
<tr>
<td>1992</td>
<td>420</td>
<td>520</td>
<td>977</td>
</tr>
<tr>
<td>1991</td>
<td>0</td>
<td>437</td>
<td>739</td>
</tr>
<tr>
<td>1990</td>
<td>0</td>
<td>255</td>
<td>508</td>
</tr>
<tr>
<td>1989</td>
<td>0</td>
<td>158</td>
<td>336</td>
</tr>
<tr>
<td>1988</td>
<td>0</td>
<td>120</td>
<td>229</td>
</tr>
<tr>
<td>1987</td>
<td>0</td>
<td>85</td>
<td>184</td>
</tr>
<tr>
<td>1986</td>
<td>0</td>
<td>79</td>
<td>190</td>
</tr>
<tr>
<td>1985</td>
<td>0</td>
<td>64</td>
<td>144</td>
</tr>
<tr>
<td>1984</td>
<td>0</td>
<td>51</td>
<td>140</td>
</tr>
<tr>
<td>1983</td>
<td>0</td>
<td>33</td>
<td>109</td>
</tr>
<tr>
<td>1982</td>
<td>0</td>
<td>26</td>
<td>91</td>
</tr>
<tr>
<td>1981</td>
<td>0</td>
<td>21</td>
<td>88</td>
</tr>
<tr>
<td>1980</td>
<td>0</td>
<td>13</td>
<td>77</td>
</tr>
<tr>
<td>1979</td>
<td>0</td>
<td>13</td>
<td>70</td>
</tr>
<tr>
<td>1978</td>
<td>0</td>
<td>10</td>
<td>71</td>
</tr>
<tr>
<td>1977</td>
<td>0</td>
<td>14</td>
<td>67</td>
</tr>
<tr>
<td>1976</td>
<td>0</td>
<td>16</td>
<td>51</td>
</tr>
<tr>
<td>1975</td>
<td>0</td>
<td>9</td>
<td>79</td>
</tr>
<tr>
<td>1974</td>
<td>0</td>
<td>6</td>
<td>53</td>
</tr>
<tr>
<td>1973</td>
<td>0</td>
<td>6</td>
<td>41</td>
</tr>
<tr>
<td>1972</td>
<td>0</td>
<td>6</td>
<td>25</td>
</tr>
<tr>
<td>1971</td>
<td>0</td>
<td>3</td>
<td>22</td>
</tr>
<tr>
<td>1970</td>
<td>0</td>
<td>8</td>
<td>24</td>
</tr>
<tr>
<td>1969</td>
<td>0</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>1968</td>
<td>0</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>1967</td>
<td>0</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>1966</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1965</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1964</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1963</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1962</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1961</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1960</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Column B also includes the term for column A, and column C also includes the terms for columns A and B. Search performed on 20 September 2001.
may be diagnosed at earlier stages in people with diabetes.\(^{12-14}\) Possibly because they are under closer medical supervision than non-diabetics. Therefore, a possible explanation for the finding is simply that wild type cases had (by chance) a more localized distribution of stage.

- It is possible that some organochlorine compounds may increase the risk of diabetes.\(^{25-26}\) If so, the overrepresentation of diabetics among the seven K-ras wild type cases would tend to mask differences in organochlorine compounds concentrations with the 17 mutated cases. This might partly explain differences with a previous study of ours: we found that mutated cases had significantly higher concentrations of organochlorines than non-mutated cases.\(^{27}\)

- The 46 subjects with a K-ras mutated tumour were almost three times more likely to be in the upper category of total coffee consumption than the 15 cases without a mutation (OR=2.78, p=0.11) (incidentally, the p value of 1.00 that the authors offer is wrong; the p value published for caffeinated coffee is also overestimated, while that for decaffeinated coffee is correct).\(^{28}\) But polyuria (frequent secretion of urine) and polydipsia (frequent thirst) are common symptoms of diabetes; increased urination is a consequence of osmotic diuresis secondary to sustained hyperglycaemia; this results in a loss of glucose as well as free water and electrolytes in the urine. Thirst is a consequence of the hiperosmolar state.\(^{29}\) It is thus plausible that the diabetic patients included in the study had a tendency to drink more liquids than non-diabetics. If so, a spuriously high presence of diabetes in the wild type group may have blurred case-control differences on coffee consumption. A slight bias in the direction we hypothesise would bring their OR closer to the figure of 3.65 that we observed based on 83 mutated and 24 wild type cases, as discussed in this journal.\(^{30}\) Should the study by Slebos et al. be thought to provide fresh additional evidence for an association between coffee consumption and K-ras mutations in exocrine pancreatic cancer? One could be inclined to think so, but the first set of questions we posed above makes the judgement uncertain. And so we may be unable to assess another potentially important gene-environment interaction.

---

**Molecular Epidemiology Can Do Better**

More needs to be known about the environmental aetiology of cancer and the causes of genetic alterations in this and other groups of diseases.\(^{31-34}\) But such knowledge is proving elusive. Perhaps for lack of appropriate studies? (For example, for lack of valid studies that care to assess oncogene-environment interactions?). Research funding is at a historical low in the United States, the nation that leads international research on cancer. The study by Slebos et al. is only the second attempt to analyse internal concentrations of organochlorine compounds and K-ras mutations in any human cancer. Beyond the merits and limitations of this and other studies, all this reflects the difficulties faced by most of us worldwide, and the limitations of some research approaches to the causes of complex diseases. An open debate on these issues should help to design, conduct, and analyse more valid and meaningful molecular epidemiological studies. It would hence help to add value to research on human genomics—modestly, but distinctively.

**Acknowledgements**

Thanks are due to J M Corominas, A Salas, L Guarnier, J Riña, S Coll, R Solá, and M Andreu for scientific advice; to D J MacFarlane, E Fernandez, J L Piñol, S Costaferda, G Castañeda, J Nog, and A Serrat for technical assistance; to A Amorós for help in drawing the figures; and to P Barbas, L Español, and O Juan for secretarial assistance.

---

**Authors’ Affiliations**

M Porto, N Malats, M Soler, L Ruiz, D Ayude, F X Real, Institut Municipal d’Investigació Mèdica, Universitat Autònoma de Barcelona and Universitat Pompeu Fabra, Barcelona, Spain

J Vioque, A Carrato, V Barberà, Universitat Miguel Hernández, Alicante, Spain

Funding: partly funded by research grants from Ministerio de Ciencia y Tecnología [CICYT SAF 2000-0097], Fondo de Investigación Sanitaria (95/0017) and Generalitat de Catalunya (BEAI 1998/400011 and DURSI 2001/SGR/406).

Conflicts of interest: none.

Correspondence: Professor M Porto, Institut Municipal d’Investigació Mèdica, Universitat Autònoma de Barcelona, Carrer del Dr Aiguader 80, E-08033 Barcelona, Spain; mporto@imim.es

Accepted for publication 3 April 2002

**REFERENCES**


6 Tomey G. Dispute as rival groups publish details of human genome [news]. BMJ 2001; 322: 381.


www.jech.com


Incomplete overlapping of biological, clinical, and environmental information in molecular epidemiological studies: a variety of causes and a cascade of consequences

M Porta, N Malats, J Vioque, A Carrato, M Soler, L Ruiz, V Barberà, D Ayude and F X Real

*J Epidemiol Community Health* 2002 56: 734-738
doi: 10.1136/jech.56.10.734

Updated information and services can be found at:
http://jech.bmj.com/content/56/10/734

**References**

This article cites 44 articles, 13 of which you can access for free at:
http://jech.bmj.com/content/56/10/734#BIBL

**Email alerting service**

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

- Epidemiologic studies (2838)
- Screening (oncology) (107)

**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/