



The European Agency for the Evaluation of Medicinal Products  
*Pre-authorisation Evaluation of Medicines for Human Use*

26 March 2002  
COMP/436/01

**COMMITTEE FOR ORPHAN MEDICINAL PRODUCTS  
(COMP)**

**POINTS TO CONSIDER ON THE  
CALCULATION AND REPORTING OF THE PREVALENCE OF A  
CONDITION FOR ORPHAN DESIGNATION**

<b>DISCUSSION IN AD-HOC GROUP OF EXPERTS ON PREVALENCE</b>	May/July 2001
<b>TRANSMISSION TO THE COMP</b>	October 2001
<b>RELEASE FOR CONSULTATION</b>	November 2001
<b>DEADLINE FOR COMMENTS</b>	February 2002
<b>ADOPTION BY THE COMP</b>	March 2002

# POINTS TO CONSIDER ON THE CALCULATION AND REPORTING OF THE PREVALENCE OF A CONDITION FOR ORPHAN DESIGNATION

## 1. INTRODUCTION

One important aim of the European Regulation on orphan medicinal products is to promote the development of products for serious rare diseases or for serious diseases where without incentives it is unlikely that the marketing would generate sufficient return to justify investment.

Where an orphan designation application is based on the claim that a condition for which the medicinal product is intended is rare, *i.e.*, that the condition affects not more than 5 in 10,000 persons in the Community, then this should be demonstrated by the sponsor using appended authoritative references.

No matter how rare a condition actually is, it is not sufficient to state that it is ‘obviously’ rare and the prevalence is far below the 5 in 10,000 limit. Generally, demonstrating that the prevalence of a condition meets the criterion will consist of a review of the literature and of any reference databases together with a critical presentation of methods, results and conclusions. Where all available data show conclusively that the population prevalence lies well-below the threshold, the fulfilment of the prevalence criterion will be a relatively simple task. In less clear situations, careful weighing of the evidence from available sources and formal numerical combinations of available data may be necessary to establish that the population prevalence lies below the threshold. It is recognised that in some very rare diseases or conditions obtaining relevant morbidity data to demonstrate that the prevalence meets the orphan criterion may be the most difficult task.

Another important point that requires clarification is the interpretation of the prevalence criterion for conditions of very short duration. For such conditions, yearly incidence rather than point prevalence will often be a more relevant measure in view of the objectives of the orphan drug legislation.

The aim of this points-to-consider document is primarily to assist the sponsor in establishing the prevalence of a condition. It suggests possible sources of data, review methods and presentation of results so that the claim can be established in a transparent and convincing way. In particular, this document addresses:

- Problem Statement and Key Definitions
- General Points to Consider
- Identification of Epidemiological Data
- Validity and Comparability of Data
- Combining Data from Different Studies
- Reporting

The document does not discuss statistical models and methods for carrying out the epidemiological studies and estimations. It is assumed that valid epidemiological designs and statistical methods are used throughout the different steps involved in the estimation and reporting of the prevalence. This document should be read in conjunction with the following regulations and guideline, which address a number of fundamental issues, such as valid definitions of a condition:

- Regulation (EC) No 141/2000 of the European Parliament and of the Council of 16 December 1999 on orphan medicinal products
- Commission Regulation (EC) No 847/2000 of 27 April 2000 laying down the provisions for implementation of the criteria for designation of a medicinal product as an orphan medicinal product and definition of the concepts ‘similar medicinal product’ and ‘clinical superiority’
- European Commission Guideline on the format and contents of applications for designation as orphan medicinal product (ENTR/6283/00)

## 2. PROBLEM STATEMENT AND KEY DEFINITIONS

**Prevalence** is traditionally defined as the number of persons with a disease or condition at a specified instant in time in a given population. It is sometimes referred to as ‘point prevalence’ and expressed as a proportion.

The ‘prevalence criterion’, which is described in article 3 (1) (a) of Regulation (EC) No 141/2000, requires the demonstration through authoritative references that the disease or condition for which the medicinal product is intended, affects not more than 5 in 10,000 persons in the Community, when the application is made. Therefore, in the context of the orphan legislation the prevalence refers to the number of persons with the condition at the time the application is made, divided by the population of the Community at that time. In the application for designation, prevalence should be expressed as the proportion of persons affected by the condition, *per 10,000*. For instance, with an estimated population in the Community of 377.6 million (as of 1 January 2001) a total of 188,800 persons correspond to a prevalence of 5 in 10,000.

For the purpose of establishing the ‘prevalence criterion’, prevalence is expressed as a proportion, and the population at risk (the denominator) should always refer to the entire population of the Community even if the population at risk of the condition is just a subset of the entire general population (e.g., ovarian cancer in women, idiopathic respiratory distress syndrome in premature newborns).

For conditions of average duration of less than one year, prevalence data should be complemented with yearly incidence data (relevant to the year of submission of the application) and the sponsor should establish that the condition affected less than 5 per 10,000 persons *during the year* when the application was submitted.

In many situations, the *true* prevalence at the time of application will not be known and the demonstration of the ‘prevalence criterion’ will be based on the *estimated* prevalence of the condition at a certain point in time. Where this is the case, there should be reasonable evidence that the estimate provided is a good approximation of the true prevalence of the claimed orphan condition in the European Union, at the time of application.

## 3. GENERAL POINTS TO CONSIDER

- The starting point for any prevalence estimation is the definition of a medically plausible condition that is generally recognised. Guidance on how to define medically plausible conditions can be found in the European Commission Guideline on the format and contents of applications for designation as orphan medicinal product (ENTR/6283/00).
- The best and most reliable sources of epidemiological data will vary depending on the condition of interest and there is no unique best source.

- If the product is intended for *prevention* or *diagnosis* of a condition, then the limit of 5 in 10,000 persons in the Community refers to persons receiving the preventive treatment or subjected to the diagnostic test and not those affected by the condition itself. More generally, if the number of persons requiring administration of a product for prevention or diagnosis of a condition exceeds the number of persons affected by the condition, then the estimation of prevalence should be based on the number of persons that are candidates for being administered the product.
- For medicinal products intended for the *treatment* of a condition, the prevalence should generally be calculated based on the number of persons affected by the condition, regardless of the number of persons who are or are not expected to receive the claimed orphan medicinal product (e.g., because of the existence of treatment methods with a better benefit/risk profile in certain sub-populations).
- Epidemiological data are more likely to be found for generally accepted conditions, defined by commonly used classification rules. For less well-defined conditions, such as subgroups of recognised conditions, data is often difficult to find or non-existent. Guidance on the medical plausibility of subsets is available in the Guideline on the format and content of applications for designation as orphan medicinal products (ENTR/6283/00). Where the sponsor claims that a subset of a condition is medically plausible, then the application should report the prevalence of the subset as well as that of the condition.
- Where comprehensive prevalence data for the Community is not available, data from individual EU Member States or national regions may still be available. Assumptions about the validity of extrapolating these data to the whole Community will have to be made and justifications on the validity of such assumptions will have to be provided. The possibility of temporal or spatial variations should be considered and appropriate adjustments should be made whenever necessary (e.g. north-south difference in the prevalence of thalassaemia or differences in patterns of hospital admission for different health-care systems). More generally, significant sources of bias for any extrapolation should be taken into account.
- When all available epidemiological sources indicate that the prevalence of a condition lies well below the limit, then the extent of precision that is required in the estimation of the prevalence is generally small. For example, a summary of main epidemiological literature or a simple merging of data from available studies may often be sufficient. Conversely, when the estimated prevalence is close to the threshold defined by the prevalence criterion, more precise evidence is generally required and this may rely on the use of complex statistical methodology, if the data allows such an approach.
- The level of detail will vary on a case by case basis, according to the availability of data and the required precision. In any case, the epidemiological section of the application should contain sufficient detail to assess the quality of the epidemiological source data, of any methods of calculations used by the sponsor and to assess the validity of the conclusions claimed by the sponsor.
- Data clearly not reflecting the number of persons affected by the condition or otherwise independent of the number of persons receiving the claimed orphan medicinal product, are not acceptable without appropriate adjustments (for example when the medicinal product is intended for treatment of *clinically manifest poisoning* in hospitalised patients, then telephone enquiry statistics from poison information centres representing *primarily suspected or minor poisonings* requiring no or minor treatment may not be relevant).

- The interdependence between prevalence, incidence and duration of the disease is well known and it follows that the definition of the duration of a condition is of particular relevance for the estimation of prevalence<sup>1</sup>. A prerequisite for any valid definition of a condition lies in its ability to capture the *entire course of the condition*. This should include, for instance, any long-term or permanent significant impairment, even if treatment-derived and even if it extends beyond the period in which pharmaceutical interventions are deemed possible or beneficial. Where the estimation of prevalence is based on particular assumptions or estimates of the average duration of the condition, then these will be subject to scrutiny as to their validity. Also, there are situations in which establishing the average duration of a condition may prove difficult. In such situations, adequate justifications should be provided in order to justify the expected duration of the condition.

#### 4. IDENTIFICATION OF EPIDEMIOLOGICAL DATA

For each condition, available information will often vary in terms of scientific value and quality of the source. Search methods should always aim to identify the most rigorous and quality sources of information.

Standard sources of information typically consist of primary epidemiological and medical literature from peer-reviewed journals and, where available, databases and registries (provided that the source of the data and methodology are documented and meet equivalent scientific standards). The strategy for identification of relevant information will generally include a search of bibliographic databases. A systematic review of all available epidemiological literature is often sufficient for producing a reliable overall estimate of prevalence. Where this is not the case then further relevant information could be identified for example through Internet searches and contact with experts.

Textbooks may be useful in pointing to relevant sources or may themselves provide useful referenced epidemiological data. However, unsubstantiated statements about the prevalence of a condition will generally be insufficient, even if derived from a textbook, monograph or a thesis on the condition.

Where standard sources may have shown to be uninformative or unreliable, the sponsor should also consider admission or discharge records of hospitals and specialised centres, surveys of General Practitioners, rare disease or patient organisations, statistics on drug use (for example the number of prescribed medications during a certain time period, reimbursement statistics) and statements from experts.

The strategy for the identification of prevalence data should be presented together with a thorough discussion of potential bias (e.g., publication bias and selection bias).

The sponsor should generally take into account information about all published studies and other accessible sources relevant to the claimed orphan condition. All source material should be adequately documented in the application, according to scientific standards. In general, where after a search of the literature the sponsor has been able to establish that the estimated prevalence is well below the threshold (for example, 1 or more orders of magnitude) and where this corresponds to general knowledge about the prevalence of a condition, then the level of detail of the information to be provided may be substantially reduced.

However, in less obvious situations, including situations where the confidence in the reliability of the data is insufficient, the validity of the overall prevalence calculation provided by the

---

<sup>1</sup> Under the assumptions of stable incidence and duration of the condition, the functional relationship between point prevalence ( $P$ ), incidence ( $I$ ) and mean duration ( $D$ ) is commonly expressed as  $P = I \times D$ .

sponsor may be seriously questioned in case additional relevant epidemiological studies are identified during the evaluation of the application. The sponsor should promptly inform the EMEA if relevant new prevalence information becomes available during the designation procedure.

## **5. VALIDITY AND COMPARABILITY OF DATA**

After all potentially relevant sources are identified exploration of sources of bias is of great importance because studies often vary considerably in terms of design, definition of the condition or methodology. Such differences may induce artificial heterogeneity that needs to be distinguished from real variation of the occurrence of the condition within the Community. These issues need to be addressed in the application, with the aim of distinguishing real heterogeneity in the occurrence of the disease from that due to differences in study characteristics.

The exclusion of individual sources or studies from the overall evaluation of prevalence can at times be justified. Exclusion criteria should be described and substantiated. Exclusions can only be justified in case of documented bias that cannot be corrected by appropriate weighting in the calculation. If studies are excluded, the validity of the obtained result should be investigated in alternative analyses based on the complete set or on different subsets of the studies.

### **5.1 Summarising the Data from Available Sources**

In many situations available data will be sufficient to demonstrate that the prevalence lies below the threshold, without the need of combining the data from different sources with the aim of producing a very precise estimate of the prevalence.

In other situations, the sponsor may choose to combine the results of relevant available sources. When there is evidence of no significant variations across studies and in the Community, this will often consist of a simple merging of data from different studies. However, when the occurrence of the condition does vary within the Community, or studies available are not representative of a defined population as a whole (because of differences in population characteristics that are determinants of the condition), the sponsor may suggest a weighted average. In this case, a transparent justification should be provided on how weights are chosen and assigned to the studies and the means by which variability among study results has been dealt with. The impact of observed or suspected major demographic differences among the individual studies should be evaluated in alternative analyses with the aim of demonstrating the validity of the combined result. Similarly, sensitivity analyses may be necessary to demonstrate the degree to which the final prevalence estimate is sensitive to the underlying assumptions or weights chosen, particularly where prevalence is suspected to be close to the designation threshold.

## Checklist for Reporting

The following key items will normally be addressed in the epidemiological section of the application:

- The strategy used for identifying prevalence data
- The strategy used for evaluating and combining available evidence
- The most important data derived from relevant sources
- Main results after combining individual studies
- A conclusion on the population prevalence made by the sponsor.

The level of detail to be reported in the epidemiological section of the application should be evaluated on a case by case basis. For instance, when all available epidemiological sources indicate that the prevalence of a condition lies well below the limit, then the extent of precision that is required in the estimation of the population prevalence is generally small. Similarly, more detail will generally be required when the prevalence of the disease or condition is close to the 5 in 10,000 limit than when it is an order of magnitude below it.

The following checklist suggests items that might be reported in the application, depending on the desired level of detail. The checklist is not to be interpreted as a requirement and it is acknowledged that for the majority of applications just the key items may be sufficient to establish the population prevalence.

### 1. CHECKLIST FOR THE REPORTING OF METHODS

#### 1.1 Definitions for the Epidemiological Assessment

Some key definitions relevant to the remainder of the epidemiology section of the application should be clearly stated. These include, for example:

- the condition of interest (addressing, if applicable, handling of associated conditions; definition of duration including recovery and handling of recurrent events);
- the population to which the product is actually expected to be administered in order to target the condition of interest;
- the chosen (calendar) time point or period for which conclusions on population prevalence are to be made.

#### 1.2 Methods for the Identification of Epidemiological Data

To facilitate a critical evaluation of the selection process, the epidemiological section of the application should include sufficient details on the search methods used and, in more general terms, on the efforts made to include all available information, such as:

- overview of the strategy for identifying relevant sources, including
  - use of bibliographic databases such as MEDLINE, search algorithm (time period included, keywords, languages included, etc.)
  - use of databases and registries
  - use of hand searching (e.g., cited references) and contacts with experts
  - methods for handling abstracts and use of unpublished material
- methods to assess the relevance or quality of information

- methods to assess bias which may affect point estimates

### **1.3 Methods for Combining the Data Identified**

In situations where the claim on the prevalence is based on a numerical combination of available studies, details about the methods for combining the data from different studies should be provided.

To facilitate, in particular, a critical evaluation of the appropriateness of the methods used for producing any quantitative summary of the data, the application should include sufficient data for the assessment of numerical methods used for combining data from different studies, if any, with a description of statistical methods (including methods for investigation of heterogeneity).

## **2. CHECKLIST FOR THE REPORTING OF RESULTS**

### **2.1 Search Results**

Search results should be reported in great detail within the body of the epidemiological section of the application and discussed in the relevant sections, if necessary, individually. In order to facilitate review, a simple overview of the results of the search should be produced in the form of tabular listings or other data summaries, (see Appendix 2 for an example of a possible format for a summary table). All identified sources should be accounted for, including the information excluded, with a justification.

Where all available studies point in the same direction of a prevalence well-below the threshold, then an exhaustive tabulation of individual details may at times be unnecessary and a justification for omitting this could be provided. In general, however, in order to facilitate the assessment of search results, the following information should be provided for relevant sources, wherever available:

- type of source
- study reference
- geographic region
- calendar years of data collection
- case definition and diagnosis procedures
- computational methods (modelling, computation of prevalence)
- most important assumptions
- study design, method of case ascertainment (and response rate, if appropriate),
- definition of the study population (including methods to establish the size of the reference population, if necessary)
- sample size
- periodic re-assessments, evidence of time trends, changes in diagnostic procedures
- reported estimated prevalence, confidence intervals and other available routine statistics (incidence, mortality, etc.)

### **2.2 Summary of Available Data and Main Results about the Population Prevalence**

The main results of the epidemiological section should be stated clearly and it is on these results that the sponsor should base the conclusive claim about the claimed population prevalence of the condition.

Typically, the main results will include the point estimate of the population prevalence and some indication of the precision of the estimation (typically a confidence interval). It is acknowledged that at times, a rigorous estimation of the prevalence will not be possible and a plausible range of hypothesised values or a worst-case scenario estimation will be the only possible summary of the available data.

If main results are derived from numerical combinations of several available sources, reporting of the following could also be considered:

- graphical/tabular summaries of individual study estimates and weight in overall estimation;
- results of analyses of heterogeneity;
- results of analyses aiming to evaluate robustness of the calculation.

### **3. REPORTING OF DISCUSSION AND CONCLUSION**

A discussion of main results and most relevant aspects related to the demonstration of the prevalence criterion should be provided, whenever appropriate. The discussion may for instance include a critical assessment of search methodology, relevance and quality of information assembled, bias, exclusion of certain sources or studies, statistical methods, justification on whether it is appropriate to combine results from different studies, main results and measures of uncertainty, validity of assumptions and the ability to generalise conclusions.

A conclusive statement should be provided on the number of persons affected by the condition in the Community at the time the designation application is made.

## Example of a Tabulated Summary of Studies

<b>(No.) Reference</b>	<b>11. BMJ 1988; 297: 1599-1602</b>	...
<b>Type of Source</b>	Literature (peer-reviewed journal)	...
<b>Region</b>	UK	...
<b>Collection Year(s)</b>	1983-1985	...
<b>Case definition</b>	Cystic fibrosis	...
<b>Data collection method</b>	Questionnaire to all consultants members of the British Paediatric Association, the British Thoracic Society and the British Association of Paediatric Surgeons (1983)	...
<b>Design</b>	Survey	...
<b>Reference population size</b>	Not provided in the publication	...
<b>Risk factors</b>		...
<b>Other</b>		...
<b>Incidence (I)</b>	one case in 2,500 live births	...
<b>Mortality (M)</b>	80% survival at 8 year-old; 50% survival at 19 year-old,	...
<b>Calculated prevalence</b>	5000 cases estimated mid-1985	...
<b>Comments</b>		...
		...