Chapter 2

Special methodological challenges when studying the diagnosis of unexplained complaints in primary care

Marloes A van Bokhoven, Hèlen Koch, Trudy van der Weijden, Geert-Jan Dinant

Journal of Clinical Epidemiology 2008;61:318-322
Chapter 2

Introduction

Unexplained complaints (UCs) can be defined as complaints that, after a proper history taking and physical examination, do not seem to be explained by either somatic or psychiatric diseases, nor by the patient’s psychosocial context. Among the UCs most frequently seen in primary care are fatigue, abdominal complaints, dizziness, musculoskeletal complaints, and headache. Thus, UCs reflect a broad range of clinical pictures. In the literature, these complaints are frequently referred to as medically unexplained symptoms, although some authors use this term for somatoform disorders as well. Therefore, we prefer the term “unexplained complaints.” Furthermore, UCs will often be self-limiting, but may also be the first sign of serious disease. These diagnostic uncertainties make physicians feel uneasy when dealing with patients presenting with UCs. This may result in unnecessary reconsulting, superfluous laboratory and other testing, or a prolonged period of waiting and seeing. In addition, the burden of disease in terms of both quality of life and costs may be high, especially when complaints become chronic.

For these reasons, a better understanding of the proper diagnostic work-up when patients present with UCs is needed. This article describes the obstacles and challenges of such a research. Based on the international standards for reporting the accuracy of diagnostic tests, we concentrate on the following issues: how to select the proper study population; how to define accurate reference standards; and how to define useful outcome measures. However, we start with some comments on how to reach a clear definition of UCs, because this influences the other issues. Several individual obstacles that are described in this commentary can be found in research on other complaints as well, but in our opinion their combination is typical for UCs. We chose the perspective of primary care to build up the article, because the prevalence of UCs is relatively high in this setting due to the early, not fully developed stage of many complaints presented. This implies that diagnostic uncertainty is high in this group.

Definition: generalizability versus reproducibility

“Unexplained” implies the subjective judgment of a complaint by the general practitioner (GP). Whether or not a GP considers a complaint unexplained depends on GP-related factors such as prior knowledge about the patient and his or her context, the information gathered during the consultation, tolerance of uncertainty, consultation skills, and the labelling of complaints as syndromes, such as irritable bowel syndrome. On the other hand, the judgment is influenced by patient factors such as the severity and pattern of the complaints and the
patients’ ability to verbalize the complaints. To generate research findings that are generalizable and relevant to daily practice, researchers thus depend on the GP to label the patient’s complaint as “unexplained”\(^8\). Inter-GP differences, recurrence patterns, and combinations of complaints might be relevant. However, researchers tend to aim for reproducibility, as subjectivity might cause bias. They might therefore want to use a strict definition and restrict inclusion to a small set of complaints and to incident cases. In addition, researchers have to decide whether to focus on UCs of recent onset or of chronic character. The latter are usually complaints for which GPs have become convinced that no somatic explanation is available. In other words, “unexplained complaint” has become a ‘psychiatric’ diagnosis instead of a diagnostic problem, for example, the diagnosis according to the Diagnostic and Statistical Manual of Mental Disorders (DSM) diagnosis of ‘somatisation disorder’. When chronic UCs no longer present diagnostic challenges, they are beyond the scope of this paper.

So, it is difficult to compare the existing research findings on UCs because researchers use different definitions. To overcome this problem, Fink et al. recently suggested the use of a classification specifically for primary care\(^9\). Others suggested multiaxial descriptions comparable to the DSM-criteria\(^10,11\). Though this may improve comparability, it does not facilitate the above-mentioned choices of in- and exclusion criteria and the duration of the complaints. In summary, the challenge for the researcher investigating UCs is to find a balance between the heterogeneity of the concept of UCs that reflects daily practice and the unequivocal definitions that are required in research.

Example 1

Schilte researched the diagnosis of mild somatisation in general practice\(^12\). A definition of somatisation frequently used for research purposes was the so-called “abridged DSM-III-R somatisation”\(^13\). However, GPs generally do not assess somatisation along a standardized set of criteria, by means of an interview or questionnaire. To translate their study findings to daily practice, Schilte et al. examined the relation between the GPs’ clinical judgments and the results on a standardized patient questionnaire\(^14\). The GPs graded the degree of somatisation (1 no somatization - 5 severe somatisation) and answered questions about problems in the communication with their patients. The symptoms of somatisation on the patients’ questionnaire were according to the abridged DSM-III-R. It appeared that somatisation according to the GPs’ clinical judgment and to the standardized measurement were weakly related to each other, but that both definitions showed acceptable construct validity. The authors recommend that for research addressing somatisation as a practical clinical problem the operationalisation of somatisation should include the clinical
Chapter 2

judgment of the practitioner. As a result of this, the focus widens from co morbid mental disturbances to communication aspects.

Selection of patients

The subjective factor in the definition of “unexplained complaints” also influences patient selection. “Unexplained” cannot be defined in a written set of in- and exclusion criteria that are applicable without involvement of the GP, for example, in medical record reviews, screening instruments, or diagnostic interviews. In addition, in the medical records GPs usually do not register the unexplained character of patients’ complaints. They note working hypotheses in very different ways, for example, symptom diagnoses, but these notations are used for explained complaints as well. It implies that GPs necessarily play a role in patient selection. This is difficult for GPs, for four reasons. First, they need to remind themselves of the research project while seeing a patient presenting with UCs. Before a GP realizes that a patient meets the inclusion criteria, he or she may already have become involved in successive stages of the consultation and may have discussed diagnostic or therapeutic plans with the patient. The moment at which the patient is eligible might then be over. In research projects that include patients presenting with complaints relating to different tracts, as may be the case in studies on UCs in general, the GPs have many but not very strong cues to remind them of the project.

The second difficulty for the GPs is that, in spite of instructions given by the researchers, it might be difficult for them to decide whether the complaint presented by the patient is “unexplained enough” to meet the selection criteria. There are differences in the definition of UCs not only between GPs but also within individual GPs, reflecting a continuum of levels of certainty. In diagnostic research, it is important to know the continuum ranging from “diseased” to “nondiseased.” It is therefore relevant to know which part of the continuum each GP considers eligible for participation. This may be assessed by means of noninclusion analysis.

The third problem might be the most difficult to deal with: the research project itself may influence the selection process. Being involved in a research project and realizing that a problem presented by a patient might remain unexplained could trigger GPs to ask a few extra questions, for instance about psychosocial influences, leading to an explanation of the complaints that the GPs would not have come up with in routine care. The same effect may be caused by knowledge of the items of the measurement instruments, especially when the diagnostics under research include signs and symptoms. It is known from other research topics that the feeling of being observed because of participation in a research project alone can influence behaviour (Hawthorne effect) although the
size of this effect is incompletely known\(^{17}\). This problem might be overcome by means of triangulation techniques, for example, interviewing GPs afterwards about the process of patient inclusion, like it is common in qualitative research\(^{18}\). The above effects of the research project itself might lead to fewer eligible patients.

When, in spite of these difficulties, a GP finds an eligible patient, the fourth problem arises. Some GPs find it difficult to tell a patient that they are unable to explain their complaints, as part of the informed consent procedure, as they might experience it as a failure that they are unable to establish a diagnosis. For patients themselves, however, the UCs are a good reason for additional diagnostics\(^{19}\).

Example 2

In a randomized clinical trial about blood test ordering for UCs in general practice (the VAgue Medical Problems In Research [VAMPIRE] trial), of which the data are currently being analyzed, the cost-effectiveness of direct test ordering is compared to a watchful waiting policy of four weeks\(^{20}\). The participating GPs were instructed to select patients presenting with one of the following five complaints: fatigue, abdominal complaints, musculoskeletal complaints, weight changes, or itch, which they considered unexplained according to the definition given in a national guideline. The number of included patients was much lower than estimated based on a previous study about the prevalence of UCs in general practice\(^{2}\). All of the inclusion difficulties described in this paragraph were mentioned by participating GPs. To check if these have not resulted in a biased patient selection, a noninclusion study is currently performed in a sample of ten participating practices, representing GPs who included many and GPs who included few patients. Records of all patients included in the VAMPIRE study in these 10 practices are compared to records of patients who presented the same complaints or were given the same working hypothesis as the study participants. Several characteristics of the patients’ complaints are compared.

Reference standard

UCs lead to a broad, heterogeneous set of possible diagnoses, although the number of diagnoses depends on the complaint definition that has been chosen. As a result, there is no single test that can be used as a reference standard. There are three possible approaches to overcome this problem\(^{21}\). The first is a diagnosis-oriented approach. In theory, one could try to use a combination of tests as a reference standard. Apart from difficulties in finding
tests that are suitable, they are easily too invasive and therefore ethically not acceptable for application, especially because the probability of finding serious pathology in patients with UCs is very low. In addition, at the moment of testing it is sometimes not clear which disease is being sought for due to the broad differential diagnosis, so it is unknown which reference standard tests should be chosen.

The second approach is based on the presence of “pathology.” In patients with UCs, it might be sufficient to make a gross distinction between the presence and absence of disease, with no further specification in a diagnosis. One method is to ask an expert panel to decide, based on test results. Another method is to follow-up the complaints for a predefined period of time and draw a conclusion afterward. A prerequisite is that the period is of sufficient length so that, if a patient has the suspected disease, this will have become manifest by the time the experts draw a conclusion. A third method is to combine both in a so-called “delayed type cross-sectional study,” which means that an expert panel draws a conclusion about the presence of pathology after a certain follow-up period. A prerequisite for the last two methods is that the “pathology” is not self-limiting during the follow-up period, because in that case no conclusion can be drawn at the end of it. When a diagnosis is established after follow-up, it might be biased by a new episode of complaints occurring during follow-up, which may be mistaken for the original episode. This is not a hypothetical risk, especially in UCs that are frequently self-limiting and can show a relapsing pattern. Both might be solved by a fourth method, in which the conclusion at the end of the follow-up period is drawn about pathology during the whole follow-up instead of at one moment. Two forms of bias remain, the first occurring when treatment during the follow-up period influences the conclusion after its termination and the second when new pathology develops so that it is not sure whether this explains the results of the test under research or reflects another situation (e.g., a patient with tiredness and a slightly elevated erythrocyte sedimentation rate who after some months is diagnosed as having Crohn’s disease).

The third approach is based on a prognostic/therapeutic consequence. This approach fits in well with daily general practice, because of the tendency among GPs to assess the implications of health problems in terms of suffering and functional impairment in daily life, instead of a diagnosis only. Tests are used to discriminate between complaints that require immediate action and complaints whose natural course can be awaited. The advantage of this approach is that it bypasses the difficult problem of drawing a diagnostic conclusion. However, this approach too might be biased by treatment that interferes with the prognosis.

The decision to adopt one of the last two approaches depends on the research question.
Example 3
Muris et al. studied the diagnostic accuracy of symptoms, signs, and test results for the diagnosis of organic gastrointestinal disease among patients with nonacute unexplained abdominal complaints\textsuperscript{23}. They did a follow-up of at least one year, during which all events were registered in the patients' records by the GPs. After follow-up, a panel of three GPs classified the diagnoses based on the international classification of problems in primary care. When the panel did not reach consensus a second panel, of professors in internal medicine and general practice, was consulted. Next, they divided all diagnoses into two categories: organic versus nonorganic. Finally, they compared neoplasms with all other diagnoses. In this way, they were able to find predictors and alarm signals of organic disease and neoplasms.

Outcome measures
As mentioned earlier, establishing a diagnosis in diagnostic research is not straightforward in patients with UCs. The options in the differential diagnosis include not only somatic diseases but also psychiatric diagnoses and complaints that become chronic, but for which no somatic explanation can be found. Naturally, outcome measures should reflect the research questions of a study. In addition to the diagnostic effects that are determined in a traditional diagnostic study, other types of outcome measures may be important as well when determining the value of a diagnostic test. One possible alternative outcome measure is costs. Because the probability of pathology in patients with UCs is low, test ordering behaviour will not have a great impact on costs of individual treatment and care. However, because the group of patients with UCs is large and many tests are requested for this group, the total costs of testing may be high. A difficulty with cost-evaluation studies is that health technology assessment methodology, similar to most diagnostic study designs, is based on singular diagnoses, while in UCs the focus may be on “pathology” in general. A second type of alternative outcome measure consists of psychosocial effects on patient and physician, for example, anxiety and satisfaction. Both GPs and patients seek methods to reduce uncertainty, and testing might influence their psychological well-being\textsuperscript{24}. A third type is that of strategic effects, for example, testing as an opportunity for not rewarding a referral request\textsuperscript{25}.

Example 4
In the VAMPIRE study, which has been described in example 2, the first primary outcome measure is the diagnostic accuracy of blood tests for serious pathology (per test and in combinations relevant for general practice) related to
signs and symptoms. The second is the effect of watchful waiting on the test characteristics. In addition, the study also pays attention to prognosis related outcome measures and costs. Outcome measures related to the GPs are, for example, level of insecurity and satisfaction with the consultation. Outcome measures related to the patient are satisfaction with care, anxiety, quality of life, utilization of health care facilities, and absence of work through illness. Related to the intervention are the costs of the training for GPs to effectively execute the watchful waiting approach.

Discussion

It may be concluded that UCs have several characteristics that distinguish them from nosologically described diseases: they are not as accurately defined, they may represent the beginning of a range of disorders, but in general practice, they are usually self-limiting, and reference standards for the diagnosis are lacking. In addition, research on these complaints is hampered by the fact that including patients is even more difficult than usual and by the fact that, though testing has limited diagnostic value, it may have “strategic” value. The combination of these characteristics makes diagnostic research into UCs not as straightforward as diagnostic research into well-defined diseases. This might be one of the reasons why such research has been sparse and adequate methodology has been lacking. Topics that deserve attention include both traditional diagnostic and nondiagnostic effects of testing, the early recognition of patients who are at risk to develop chronic complaints, the effects of a watchful waiting strategy on the course of complaints, and implementation of new findings in daily practice. The further development of research methodology for this topic is a challenge in its own right.
References


