

BMJ Open Gut Feelings Questionnaire in daily practice: a feasibility study using a mixed-methods approach in three European countries

Marie Barais,¹ Margje W J van de Wiel,² Nicolas Groell,¹ Antoine Dany,³ Tristan Montier,⁴ Paul Van Royen,⁵ Erik C F Stolper^{5,6}

To cite: Barais M, van de Wiel MWJ, Groell N, *et al.* Gut Feelings Questionnaire in daily practice: a feasibility study using a mixed-methods approach in three European countries. *BMJ Open* 2018;**8**:e023488. doi:10.1136/bmjopen-2018-023488

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2018-023488>).

Received 9 April 2018
Revised 27 September 2018
Accepted 4 October 2018



© Author(s) (or their employer(s)) 2018. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Marie Barais;
marie.barais@gmail.com

ABSTRACT

Objectives The validated Gut Feelings Questionnaire (GFQ) is a 10-item questionnaire based on the definitions of the sense of alarm and the sense of reassurance. The purpose of the GFQ is to determine the presence or absence of gut feelings in the diagnostic reasoning of general practitioners (GPs). The aim was to test the GFQ on GPs, in real practice settings, to check whether any changes were needed to improve feasibility, and to calculate the prevalence of the GPs' sense of alarm and sense of reassurance in three different countries.

Setting Primary care, six participating centres in Belgium, France and the Netherlands.

Participants We performed a think-aloud study with 24 experienced Dutch GPs, GP trainees and medical clerks who filled in the GFQ after diagnosing each of six case vignettes. We then performed a feasibility study in two phases, using a mixed-method approach, with 42 French and Dutch GPs in the first phase and then 10 Belgian, 10 Dutch and 10 French GPs in the second phase. All GPs filled in the GFQ after each of eight consultations with patients presenting new complaints and were subsequently interviewed about the use of the GFQ.

Outcome measures GPs' experiences on using the GFQ in real practice, more specifically the average time needed for filling in the questionnaire. The prevalence of GPs' sense of alarm and sense of reassurance.

Results The modified version of the GFQ, created without altering the sense of the validated items, was easy to use in daily practice. The prevalence of the GPs' sense of alarm occurred during 23%–31% of the included consultations.

Conclusions After a two-step study and several minor adaptations, the final version of the GFQ proved to be a feasible and practical tool to be used for prospective observational studies in daily practice.

INTRODUCTION

The gut feelings (GFs) which may arise during the process of diagnostic reasoning by general practitioners (GPs) have been defined as a sense of alarm and a sense of reassurance.¹ The sense of alarm is an uncomfortable feeling, experienced by the GP, that something does not fit in a patient's clinical

Strengths and limitations of this study

- Testing the use of a questionnaire such as the Gut Feelings Questionnaire (GFQ) in two different settings (think-aloud in an experimental environment first, and then during office hours, in three different healthcare systems) was quite unique.
- The GFQ is directly derived from the consensual definition of gut feelings: its added value is the detailed and precise way it measures general practitioners (GPs) gut feelings.
- Quite a number of the GPs did not fill in the questionnaire right after the consultation but completed it later that day.

presentation although he/she has not (yet) found specific indications. The sense of reassurance means that a GP feels secure about the health status of the patient, even if he/she is not certain about the diagnosis. The sense of alarm activates the diagnostic process and initiates specific management to prevent serious health problems.¹ GFs play an important role in the diagnostic reasoning process of GPs helping them to navigate in the complex and uncertain diagnostic situations encountered in practice.² It has been described as a third track, alongside medical decision-making and medical problem solving, enabling the physician to commute between non-analytical and analytical reasoning processes.²

In studies earlier done, the sense of alarm and the sense of reassurance were defined following a qualitative analysis of the text of several focus groups on the topic and a subsequent Delphi consensus procedure.^{1 3} The items of a Dutch Gut Feelings Questionnaire (GFQ) were based on these definition criteria. The objectives of the questionnaire are to determine the presence or absence of GFs in GPs' diagnostic reasoning at the end

of a consultation based on a clear consistent definition of the concept. This questionnaire measures whether a GF is present (ie, not just by a yes or no response, as is mostly done in clinical studies about GFs⁴⁻⁶) and differentiates between the sense of reassurance and the sense of alarm by more precise statements reflecting the outcomes of the diagnostic reasoning process. The GFQ was validated by a construct validation procedure using case vignettes.⁷ A principal component analysis (PCA) showed one component explaining 70.2% of the total variance with the sense of alarm and the sense of reassurance as opposites. The internal consistency of the GFQ proved to be high (Cronbach's alpha=0.91). The kappa with quadratic weighting was substantial (0.62, 95% CI 0.55 to 0.69).⁷ A linguistic validation procedure was performed to obtain an English version of the questionnaire⁷ (figure 1).

An international network group called COGITA was established with the aim of coordinating and stimulating research into the role of GFs in general practice (see www.gutfeelingsingeneralpractice.eu). Linguistic validation procedures produced a French, Polish and German version of the GFQ,⁸ and a Spanish and a Catalan version⁹ (publication in process).

The GFQ can be used in studies measuring the prevalence of GFs and their predictive value for a serious disease.¹⁰ The questionnaire was, however, never evaluated by GPs in real settings during office hours. The aim of this study was to explore the practicability of the GFQ, that is, feasibility and acceptability as experienced by GPs when using the instrument in daily practice, and to calculate the prevalence of the GPs' sense of alarm and sense of reassurance in three different countries.

We conducted a think-aloud study to explore whether the way experienced GPs, GP trainees and medical students understood the GFQ items was in line with what we aimed for when composing the questionnaire. The next step was a feasibility study in daily practice with the original GFQ. By collecting quantitative data, we measured and compared the prevalence of GPs' GFs in different countries. As these two phases led to some adaptations, the modified questionnaire was retested during a second feasibility study. Figure 2 shows the steps taken to test the original version of the GFQ and to arrive at the final version as the result of both studies.

METHODS

Think-aloud study

Participants

Three groups, differing in their level of experience in Dutch general practice, participated in a think-aloud study on diagnostic reasoning. Participants were eight experienced GPs (seven female; average experience in GP practice was 18.6 years, ranging from 6 to 29 years), eight first-year GP trainees (five female; average clinical experience before their traineeship was 24.5 months, ranging from 9 to 53 months) and eight advanced medical students (seven female) doing their internship in general

practice at Maastricht University. The experienced GPs were recruited through a snowball strategy in the Netherlands, whereas the trainees and medical students were approached via the Department of Family Medicine at Maastricht University in the Netherlands.

Materials

Six case vignettes were developed based on actual accounts from patients. Each case vignette briefly described the complaints, medical history and results from history taking and physical examination. The cases described patients with myalgia, asthma, cardiomyopathy, pancreatic carcinoma, panic disorder and pulmonary embolism. In the real-life situation, three cases had produced a sense of alarm and three a sense of reassurance. Four cases had previously been used in the validation study.⁷ The original Dutch GFQ was used (figure 1).

Procedure

Participants were asked to diagnose each case while thinking aloud, and to fill in the GFQ afterwards, still thinking aloud. They were reminded to think aloud if they were silent for more than 5 s. The session took place in the GPs' offices or in a room at the university and lasted 30–80 min. All participants received a small gift at the end of the session.

Data analysis

All think-aloud protocols collected while participants filled in the GFQ were transcribed verbatim and analysed. We performed a thematic content analysis to summarise, per item, the problems participants encountered in interpreting and responding to the items in the GFQ.¹¹

Patient and public involvement

Patients were not involved.

Feasibility study 1

Participants

The participating GPs were recruited using a purposive sampling strategy according to criteria which could influence decision making: age, gender, location of practice (rural area means under 5000 inhabitants). Twenty French GPs from Brittany and 22 Dutch GPs, including 23 males and 19 females, aged from 28 to 64 years old, from different areas (33 urban and 9 rural areas), participated in this study. They were not given any financial incentive to take part.

Materials and procedures

A mixed-methods approach was chosen. The GPs were instructed to fill in the GFQ, during their office hours, for 8 days in a 2-week period. They were asked to include only the first consultation of the day with an adult patient, aged over 18 years, with a new reason for a consultation. After completing the eight questionnaires, the participating GPs were asked to estimate the time they needed to fill in the GFQ in minutes and were interviewed at their office or by phone. The interview guide was composed

BE-version, real practice

	Completely Disagree	Disagree	Neutral	Agree	Completely Agree
1. I feel confident about my management plan and/or about the outcome: it all adds up.	0	0	0	0	0
2. I am concerned about this patient's state of health: something does not add up here.	0	0	0	0	0
3. In this particular case, I will formulate provisional hypotheses with potentially serious outcomes and weigh them against each other.	0	0	0	0	0
4. I have an uneasy feeling because I am worried about potentially unfavourable outcomes.	0	0	0	0	0
5. This case requires specific management to prevent any further serious health problems.	0	0	0	0	0
6. What course of action have you chosen? (Please tick one answer.) I will ... <input type="checkbox"/> Wait and see. <input type="checkbox"/> Not yet take action, but will invite the patient for a follow-up appointment either face-to-face or by phone. <input type="checkbox"/> Arrange further testing (laboratory tests, X-rays, etc.). <input type="checkbox"/> Arrange further testing, and in the meantime, I will start treatment (medicinal or other). <input type="checkbox"/> Start treatment, but will not arrange a follow-up. <input type="checkbox"/> Start treatment and will invite the patient for a follow-up appointment either face-to-face or by phone. <input type="checkbox"/> Refer the patient.					
7. This patient's situation gives me reason to arrange a follow-up visit sooner than usual or to refer him or her more quickly than usual to a specialist. <input type="checkbox"/> Yes <input type="checkbox"/> No					
8a. What do you consider to be the most likely diagnosis? (Please tick one answer.) <input type="checkbox"/> My most likely diagnosis is <input type="checkbox"/> There are several possible diagnoses; I am unable to choose one at this moment.					
8b. And which diagnosis will determine your management?					
9. How confident are you in the diagnosis that you indicated under 8b as determining your management? ____%					
10. Please indicate what kind of gut feeling you have at the end of the consultation: <input type="checkbox"/> Something is wrong with this picture. <input type="checkbox"/> Everything fits. <input type="checkbox"/> Impossible to say, or not applicable.					

Figure 1 Original version of the GFQ. GFQ, Gut Feelings Questionnaire.

of two open-ended questions, which aimed to explore the experience with the questionnaire in more depth: ‘What do you think about the questionnaire’s integration into your daily practice?’ and ‘Which elements should be improved following your experience of filling in 8 questionnaires?’. Most interviews were held within 2 weeks of the 2-week period and were audio recorded

and transcribed verbatim. The interviews lasted between 3 and 18 min.

Data analysis

The analysis of the qualitative data was descriptive, using a thematic content analysis. The French and Dutch researchers coded the transcripts in an independent and

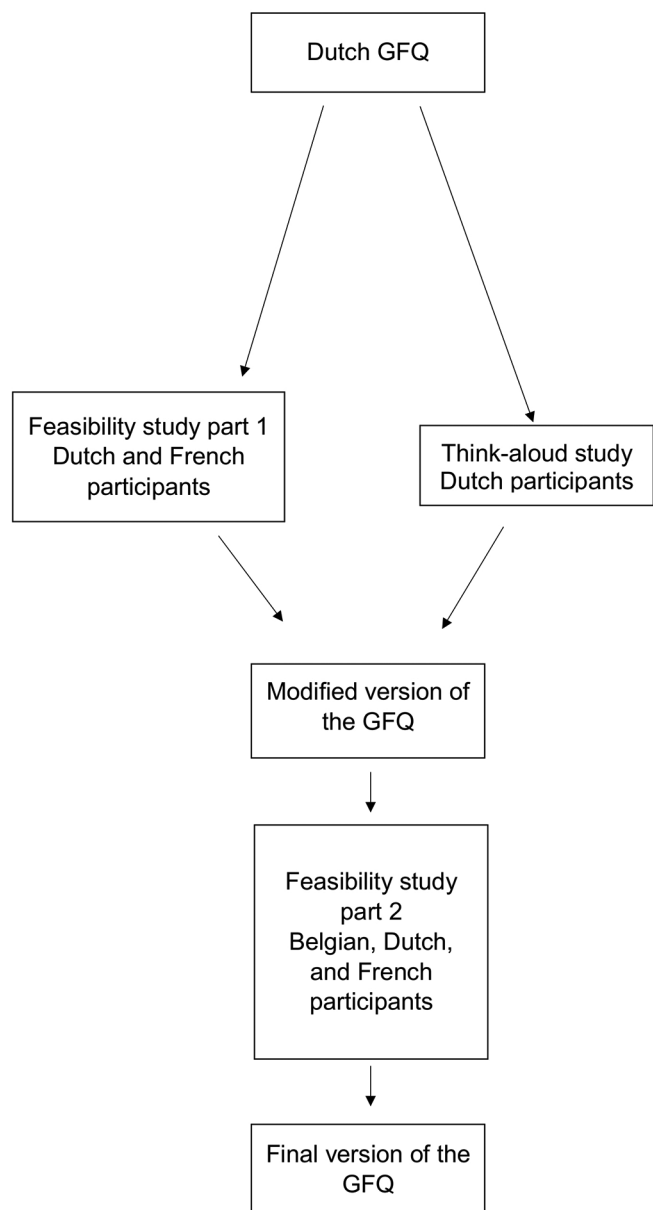


Figure 2 Scheme of the study. GFQ, Gut Feelings Questionnaire.

open way, categorised their codes and established codebooks. After having reached consensus, they merged their codebooks, adapted the codes in the different texts and reanalysed the texts. Finally, they selected the most appropriate quotes to illustrate each code in each language.¹¹ QSR NVivo V.11.0 software was used to perform the analysis.

The quantitative data, that is, the answers to items 1–5 and 10 (figure 1) were analysed with a χ^2 test using specific criteria. A sense of alarm was considered as present when the answer to item 10 indicated a sense of alarm or when the answer to item 10 indicated that it was not applicable and at least one of the scores for items 2–5 was higher than 3/5. A sense of reassurance was considered as present when the answer to item 10 indicated a sense of reassurance or when the answer to item 10 indicated that

it was not applicable and the score for item 1 was higher than 3/5. GFs were considered absent when the answer to item 10 indicated that it was not applicable and none of the scores for items 2–5 was higher than 3/5 and the score for item 1 was lower than 4/5. These cut-off criteria were chosen in line with the study protocol of the study on the accuracy of the sense of alarm when faced with chest pain and dyspnoea.¹⁰

Feasibility study 2

Participants

The modified GFQ was tested in real practice in Belgium, the Netherlands and France, with 10 GPs from each country. The participating GPs were recruited using the same purposeful sampling strategy as in feasibility study 1. Ten Belgian GPs, 10 Dutch GPs and 10 French GPs from Brittany, 15 males and 15 females, aged from 27 to 65 years, from different areas (26 urban and 4 rural areas) participated in the study. The participants were not incentivised to take part.

Materials and procedure

We used the same procedures as in the first feasibility study but presented the participants with the modified GFQ (figure 3). Most interviews were held within 2 weeks of the 2-week period and lasted 5–30 min.

Data analysis

We conducted the same thematic content analysis, using the codebook from the first feasibility study. The quantitative data, that is, the answers to items 1–6 and 11 were analysed with a χ^2 test.

RESULTS

Think-aloud study

The analysis of the think-aloud protocols revealed that some participants interpreted four of the GFQ items in a slightly different way than we intended. There were no systematic differences between the three groups. Based on these findings, we suggested small adaptations to the phrasing of two items and to the order of items to avoid misunderstanding.

Regarding item 1: ‘I feel confident about my management plan and/or about the outcome: it all adds up’. Many participants were confused by the two elements of the question, that is, management plan and/or the outcome. An experienced GP, for example, said: ‘I feel confident about the management I have in mind, but there’s something wrong... It’s a strange case’ (GP no 24). The focus for this item is ‘adding up’, so we suggested a reversal of the wording of this first item: ‘It all adds up. I feel confident about my management plan and/or about the outcome’.

Regarding item 3: ‘In this particular case, I will formulate provisional hypotheses with potentially serious outcomes and weigh them against each other’. Several participants found this criterion stated the obvious. An experienced GP said: ‘Yeah, of course, you always do that

Gut Feelings Questionnaire	Completely disagree	Disagree	Neutral	Agree	Completely agree
	1	2	3	4	5
1. Please indicate what kind of gut feeling you have at the end of the consultation. If you cannot answer this question now, go further to the next 9 items and give your answer afterwards. <i>○ something is wrong with this picture.</i> <i>○ everything fits.</i> <i>○ impossible to say, or not applicable.</i>					
2. <i>It all adds up.</i> I feel confident about my management plan and/or about the outcome.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. <i>Something does not add up here.</i> I am concerned about this patient's state of health.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. In this particular case, I will formulate provisional hypotheses with potentially serious outcomes and weigh them against each other.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. I have an uneasy feeling because I am worried about potentially unfavourable outcomes.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. <i>To prevent any further serious health problems</i> this case requires specific management	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. <i>This patient's situation gives me reason to arrange a follow-up visit sooner than usual or to refer him or her more quickly than usual to a specialist.</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. Which diagnosis/es are you thinking about?					
9. What management have you chosen? (Please tick one answer.) I will... <input type="radio"/> <i>not yet take action, wait and see.</i> <input type="radio"/> <i>not yet take action, but advise the patient to come back if the problem persists.</i> <input type="radio"/> <i>not yet take action, but invite the patient for a follow-up appointment either face-to-face or by phone.</i> <input type="radio"/> <i>order further testing (laboratory tests, X-rays, etc.).</i> <input type="radio"/> <i>order further testing, and in the meantime, I will start treatment (medicinal or other).</i> <input type="radio"/> <i>start treatment, but will not arrange a follow-up.</i> <input type="radio"/> <i>start treatment and give the advice to the patient to come back if the problem persists.</i> <input type="radio"/> <i>start treatment and invite the patient for a follow-up appointment either face-to-face or by phone.</i> <input type="radio"/> <i>refer the patient.</i>					
10. Which diagnosis has determined your management?					
11. <i>This question is the same as question 1. If you have already given an answer, there is no need to answer this question again.</i> Please indicate what kind of gut feeling you have at the end of the consultation: <input type="radio"/> <i>something is wrong with this picture.</i> <input type="radio"/> <i>everything fits.</i> <input type="radio"/> <i>impossible to say, or not applicable.</i>					

Real practice version 6 June 2016
 © 2016 COGITA ALL RIGHTS RESERVED

Figure 3 Modified version of the Gut Feelings Questionnaire after the think-aloud study and the first feasibility study with additions or changes from the first version shown in italics.

in clinical reasoning' (GP no 23). Although this remark may show that the item does not actually discriminate, we decided to leave it in the questionnaire as it was one of

the statements agreed on in the consensus procedure¹ and it also fitted in with the other items in the construct and consistency validation procedures.⁷

Regarding item 5: 'This case requires specific management to prevent any further serious health problems', Many participants answered 'yes', even if they had a sense of reassurance, due to safety netting or watchful waiting. However, this item defines the sense of alarm in the consensus definitions.¹ To emphasise the prevention of serious health problems, we suggested a reversal of the wording of this item, modifying to read: 'To prevent any further serious health problems this case requires specific management'.

Regarding item 10: 'Please indicate what kind of GF you had at the end of the consultation', several participants gave the impression that their answers to this last item were the conclusion of a rational reasoning process, based on a series of logical arguments built up from the previous items, to find an answer to item 10, rather than indicating their experience of a GF. The ranking of the first nine items could induce a bias in the answer to the last one. To avoid this, we proposed starting with the item about GFs, thereby moving item 10 to the top of the list. We also proposed repeating this item at the end of the list for those participants who were not able to answer this question at the beginning. It was only for those participants who did not answer item 1 that we used item 11 as the indication of the presence or absence of a GF. We suggested changing the order of the items in line with the usual steps of the diagnostic process.

Feasibility study 1

Qualitative results

The interviews with GPs showed some important issues regarding the feasibility of the GFQ. The GPs mainly commented on deciding how to complete the questionnaire and also commented on some of the items. There were no differences between the comments of the French and Dutch participants. They encountered the same difficulties and misunderstandings. We summarise the main findings below and illustrate these with quotes.

The GPs were asked to fill in the GFQ after each consultation in daily practice. They needed to take some extra time to do this and most succeeded, but several GPs could not deal with it immediately after the consultation and postponed it until a more suitable moment.

Of course, we're used to just following the routine, and it did very much interrupt the routine. But it took very little time. I was filling in how much time it took, and well, 2 min. (Dutch GP no 4)

Some GPs responded about the timing needed to fill in the GFQ. At the end of a block of consultations, mostly. I occasionally did one or two immediately after the consultation if I had a gap in the flow of patients, as I had some time available and it was a new complaint, but I often did it at the end of the morning or the day, thinking 'Oh I must have seen some new people and I need to fill in the questionnaires'. (Dutch GP no 2)

I did not fill in the questionnaire right after the consultation, I preferred to do it at the end of the day, because in fact, technically, during my consultation, I don't have the time to do it, well I am being honest, aren't I?. (French GP no 2)

Some participants experienced problems in answering items 6, 8 and 9. In item 6, participants were asked about their management. Some participants found the list with possible courses of action incomplete and gave suggestions for improvement.

I tended to think 'In some cases I can examine the patient and start therapy and make a follow-up appointment at the same time. So, I mean, I couldn't fit it all in one line. (Dutch GP no 10)

In item 8, they were asked to provide the most likely diagnosis and the diagnosis that determined their management. Several participants said that they were confused about the difference between these two types of diagnoses and suggested how to improve the clarity of the question. In item 9, they were asked to indicate how confident they were (as a percentage) about their management determining diagnosis. This question also raised confusion.

The only downside for me, was the worry of differentiating between the questions, between question number 8A and 8B, between the diagnosis and the hypothesis, I often put the same answer in both boxes. (French GP no 5)

For my own line of reasoning I would have preferred an extra question inserted here, like: what options are you thinking of? Differential diagnoses 1, 2 and 3. And which one do you consider to be the most important one, the one you absolutely want to exclude? And which one would you perhaps want to address?. (Dutch GP no 4)

So I can argue about that, but I find it more difficult to express it in a number, medically speaking... [...] you'd say 'which diagnosis would determine your management? and that would be pneumonia. But you think the likelihood is very small: 5%. [...] so you wouldn't have an X-ray done or do a CRP test... and then under 8b it says "which diagnosis would determine your management?" You write down pneumonia, but that's not really true, is it? Because your management is not aiming to exclude pneumonia.... (Dutch GP no 6)

Based on these comments, we proposed modifying the two diagnostic workup items (items 6 and 9): we added more options for the course of action in item 6 and we decided to remove item 9 where participants had to assess their confidence in their policy determining diagnosis in terms of a percentage. The seven validated items concerned with GFs were retained.

The changes in the GFQ we proposed, based on the think-aloud study and the first feasibility study, were discussed during two consensus meetings of COGITA

researchers (Marburg 2015, Tel Aviv 2016) (<http://www.gutfeelings.eu/list/cogita-expert/>). Afterwards, based on our concerted efforts, we formulated a modified version of the GFQ, did a linguistic validation of the new elements and changed the presentation of the questionnaire into a more visual and ergonomic format (figure 3).

Quantitative results

Out of the 348 questionnaires collected during this second phase (8–10 per GP), 336 were analysable, 12 were non-analysable because of missing data. In total, 77 (23%) were concerned with a sense of alarm, 242 (72%) with a sense of reassurance and there were 17 (5%) where no GF was applicable. The internal consistency was high (Cronbach's $\alpha=0.88$). A PCA showed one component explaining 68.6% of the total variance, with the sense of alarm and the sense of reassurance as opposites. There were no significant differences between the Dutch and French GPs. They expressed the same prevalence of a sense alarm and the same prevalence of a sense of reassurance. The median average time estimated by GPs for filling in the GFQ was 1 min for the Dutch GPs and 2 min for the French GPs without significant difference between the Dutch and the French GPs.

Feasibility study 2

Qualitative results

The results of the analysis made clear that GPs had no major problems filling in the modified GFQ. The comments of the Belgian, Dutch and French participants did not differ. The practicability was good and using the GFQ took only a small amount of time.

I'd fill it in, and then it was really just a quick question-and-answer process, so it was easy going. (French GP no 22)

I don't remember having problems or saying to myself « it doesn't work [...] but honestly, I actually had a feeling of fluidity. (French GP no 24)

The GPs did not consider it a burden.

First consultation of each day, yeah, it was easy, the questions were precise enough so that it did not take three hours from the middle or at the beginning of a general practice consultation. (French GP no 21)

As in the first feasibility study, several GPs did not fill in the GFQ right after the consultation but at the end of the office hours or of the day. They did not want to interrupt the sequence of consultations with the questionnaire. They did not report recall difficulties when answering the questionnaire.

And there are some, I guess about half of them, I filled in immediately after the consultation; the rest were done in the evening, when I get to the end of the list of patients and thus fill in the register, so I filled in some of them, but the majority, more than

half of them, were completed just after the consultation. (Belgian GP no 8)

Two Dutch GPs stopped filling in the GFQ after the first question and misunderstood the formulation of this first item: 'Please indicate what kind of GF you have at the end of the consultation. If you cannot answer this question now, please answer the following nine questions, then give your answer to question 1, which is repeated at the end of the questionnaire'. They did not reply to the next nine items (no 2 to no 10).

Some participants stressed the role of the instructions before filling in the questionnaire for the first time. They highlighted the distinction between GFs in their own decision-making process and feelings of empathy towards a patient regarding a bad prognosis.

I can have an uneasy feeling—when it all fits. For example, I had a man with haematuria and no dysuria, no pollakiuria, a smoker. I thought, 'this is wrong, it's all about bladder or kidney cancer'. At the same time I thought, 'It's all right, I feel comfortable with the further approach I have in mind... I feel comfortable that the story is clear, namely, that it's very straightforward; I also know what to do now, but I know too that the outcome will not be very good. (Dutch GP no 30)

Item 8 was a bit confusing for some participants. They did not understand that this item asked for the first three diagnoses that came into their minds and mentioned the same diagnosis in both items 8 and item 10.

It was the question 'Which diagnosis determined your course of action?' Yeah well, I found that... it may be just me who did not understand the fact that it was potentially in the plural for question 8, but I had the impression of an overlap of questions 8 and 10 because, 'Which diagnosis/diagnoses are you thinking about?' obviously includes the diagnosis which determines my course of action. (French GP no 22)

Some participants also described how their GFs arose or disappeared during a consultation. They would have had some space in the questionnaire to describe their diagnostic reasoning process.

So we just have some cognitive dissonance at the time, which is to say that, in the evening, we start thinking 'perhaps I should have done something different' or 'you feel at ease, you close that file and move on to something else'. (French GP no 26)

Based on this finding, we added a sentence at the end of the questionnaire, allowing the participants to share some thoughts about their diagnostic reasoning.

Quantitative results

Out of the 263 questionnaires collected during this second phase (8–11 GFQs per GP), 259 were analysable, 7 were non-analysable because of missing data. Eighty-two (31%) were concerned with a sense of alarm and 177

(69%) with a sense of reassurance felt by the GPs. There was no significant difference between the Belgian, Dutch and French GPs' answers. They expressed the same prevalence of the sense of alarm and the same prevalence of the sense of reassurance. The median average time estimated by GPs for filling in the GFQ was 2 min for the Belgian, Dutch and French GPs. A PCA confirmed unidimensionality with one component explaining 72.3% of the total variance. The internal consistency was high (Cronbach's $\alpha=0.90$).

We compared the prevalence of the sense of alarm and the sense of reassurance in both feasibility studies with a χ^2 test. There was no significant difference between the two samples.

After the second feasibility study, we added some minor changes to the items 1, 8 and 11. We rephrased item 1, adding 'If you cannot answer this question now, please answer the following nine questions, then give your answer to question 1, which is repeated at the end of the questionnaire'. We wrote both singular and plural forms of 'diagnosis' in item 8, preferring the formulation 'have in mind' instead of 'thinking about' and suggested 'max. 3': 'What diagnoses (or diagnosis) do you have in mind? (max. 3)'. We added the following sentence after item 11: 'If you want to share some thoughts about your diagnostic reasoning, please use the back of this questionnaire' (figure 4). We also agreed on the instructions prior to filling in the questionnaire. In these instructions, we explain how items 2–7 are derived from the definitions of the sense of alarm and the sense of reassurance and how to fill in item 1, item 8 and item 10. In order to minimise rationalisations afterwards, we also emphasised to immediately fill in the questionnaire to grasp GPs' experience during the diagnostic process (preventing recall bias) for each patient who needs to be included in the study (preventing selection bias). The instructions should be embedded within the context and aim of any future study¹¹ (see box 1). In this particular study, we specified to fill in the questionnaire for the first consultation of the day with an adult patient, aged over 18 years, with a new reason for a consultation.

DISCUSSION

Through a two-step study, we evaluated the feasibility and practicability of the GFQ in real practice. The main objective of this questionnaire was to determine the presence or absence of GFs in GPs' diagnostic reasoning and to differentiate between the sense of reassurance and the sense of alarm by precise statements which reflect the outcomes of the diagnostic reasoning process. The first step, a think-aloud study and a feasibility study, led to small modifications concerning the order of items and to some small adaptations of the wording of two items. The modified version of the GFQ was created without altering the sense of the seven validated items. The second step, a repetition of the feasibility study but with the modified questionnaire, led to minor changes. The prevalence of

GFs in the two phases of the feasibility study was similar in Belgium, France and the Netherlands showing that GPs experienced a sense of alarm in 23%–31% of the reported cases.

Strengths and weaknesses of the study

Seventy GPs from Belgium, France and the Netherlands were involved in the evaluation of the questionnaire in real settings. The same misunderstandings and difficulties in filling in the questionnaire occurred in all three countries. In addition, a similar prevalence of both the sense of alarm and the sense of reassurance was found in all three countries. Even though French GPs do not have an idiomatic expression for GFs, unlike Dutch and Belgium GPs ('plus/niet-plus'), the GFQ measures their sense of alarm and sense of reassurance in the same way.¹² The linguistic validation procedure used to translate the GFQ from Dutch to English and then from English to French has been found to guarantee the cultural transposition from Dutch to French.⁸ In spite of the differences between healthcare systems, the French and the Dutch versions of the GFQ do examine the same phenomenon. The GFQ is also feasible across practice settings in different countries. The internal consistency of the original Dutch language GFQ was high (Cronbach's α 0.90) as shown in the validation study⁷ and continued to be high in the two cross-border feasibility studies (respectively, 0.88 and 0.90). The outcomes of the factor analysis in both feasibility studies were similar to the original validation study. Our studies reaffirmed the transculturality of the GFs concept.^{12 13}

Testing the use of a questionnaire such as the GFQ in two different settings (think-aloud in an experimental environment first, and then during office hours, in three different healthcare systems) was, as far as we know, quite unique. However, it enabled us to adapt the questionnaire in response to the participating Belgian, Dutch and French GPs' opinions and pragmatic concerns.

We started with the item asking for the presence of a GF in diagnostic reasoning to capture their experience immediately after the consultation: the first item is now 'Please indicate what kind of GF you have at the end of the consultation'. We repeated this item at the end of the questionnaire for those participants who were not able to answer this question at the beginning. It was only for participants who did not answer item 1 that we used item 11 as the indication of the presence or absence of a GF. There might be a risk that the last group will also use their analytical reasoning in finding an answer to item 11 but, in any case, we reduced that risk by also putting the question at the top of the questionnaire. To minimise rationalisations afterwards, we emphasised in the instructions to immediately fill in the questionnaire to better grasp GPs' experience during the diagnostic process.

The prevalence of the sense of alarm seemed to be higher in the second feasibility study than in the first one (23% vs 31%) but statistically there is no difference. Both studies took place in winter, with the same incidence of

Gut Feelings Questionnaire	Completely disagree	Disagree	Neutral	Agree	Completely agree
	1	2	3	4	5
1. Please indicate what kind of gut feeling you have at the end of the consultation. If you cannot answer this question now, <i>please answer the following nine questions, then give your answer to question 1, which is repeated at the end of the questionnaire.</i> <input type="radio"/> something is wrong with this picture. <input type="radio"/> everything fits. <input type="radio"/> impossible to say, or not applicable.					
2. It all adds up. I feel confident about my management plan and/or about the outcome.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. Something does not add up here. I am concerned about this patient's state of health.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. In this particular case, I will formulate provisional hypotheses with potentially serious outcomes and weigh them against each other.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. I have an uneasy feeling because I am worried about potentially unfavourable outcomes.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. To prevent any (further) serious health problems requires specific management of this case.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. This patient's situation gives me reason to arrange a follow-up visit sooner than usual or to refer him or her more quickly than usual to a specialist.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. <i>What diagnoses (or diagnosis) do you have in mind? (max. 3)</i>					
9. What management have you chosen? (Please tick one answer.) I will... <input type="radio"/> not yet take action, wait and see. <input type="radio"/> not yet take action, but advise the patient to come back if the problem persists. <input type="radio"/> not yet take action, but invite the patient for a follow-up appointment either face-to-face or by phone. <input type="radio"/> order further testing (laboratory tests, X-rays, etc.). <input type="radio"/> order further testing, and in the meantime, I will start treatment (medicinal or other). <input type="radio"/> start treatment, but will not arrange a follow-up. <input type="radio"/> start treatment and give the advice to the patient to come back if the problem persists. <input type="radio"/> start treatment and invite the patient for a follow-up appointment either face-to-face or by phone. <input type="radio"/> refer the patient.					
10. Which diagnosis has determined your management?					
11. This question is the same as question 1. If you have already given an answer, there is no need to answer this question again. Please indicate what kind of gut feeling you have at the end of the consultation: <input type="radio"/> something is wrong with this picture. <input type="radio"/> everything fits. <input type="radio"/> impossible to say, or not applicable. <i>If you want to share some thoughts about your diagnostic reasoning, please use the back of this questionnaire.</i>					

Real practice version 21 May 2017
© 2017 COGITA ALL RIGHTS RESERVED

Figure 4 Final version of the Gut Feelings Questionnaire with additions or changes from the modified version shown in italics.

diseases. It can be an accidental finding which is confirmed by the fact that there is no statistical difference. Further studies with the GFQ in clinical practice are needed to examine the prevalence of GFs in general practice and its predictive validity in different contexts.

The questionnaire was modified after the two phases of the study. Now we have a questionnaire formatted by GPs, for GPs, working in three European countries. A few weeks after the start of the studies, 600 questionnaires had already been included which is remarkable

Box 1 Instructions before filling in

The purpose of the questionnaire is to determine the presence or absence of gut feelings in diagnostic reasoning. These gut feelings are defined as a sense of alarm and a sense of reassurance. A 'sense of alarm' implies that a general practitioner worries about a patient's health status, even though he/she has found no specific indications yet; it is a sense that 'there's something wrong here'. A 'sense of reassurance' means that a GP feels secure about the further management and course of a patient's problem, even though he/she may not be certain about the diagnosis: everything fits in. The items 2–7 of the questionnaire are derived from these definitions. In item 8, you will be asked to suggest a maximum of three diagnoses you have in mind concerning the patient. In item 10, you will have to write which diagnosis you used to determine your course of action. In order to avoid selection bias and to reflect your experience during the diagnostic process, we urgently ask you to fill in the questionnaire for each patient who needs to be included in the study directly after the consultation. Please, read the questionnaire, so we can discuss any questions you might have.

and might indicate how practical the questionnaire is in daily practice. Including research while practising is quite unusual for GPs.¹⁴ Lack of time is usually given as the major cause of limited GP availability other than for patient care.¹⁵ Quite a number of the GPs failed to follow the full instructions given prior to the feasibility studies. It did not always appear to be feasible to fill in a questionnaire right after a consultation. These GPs mentioned, however, that when responding to all the items, they were able to recapitulate the information regarding the patients involved without any problems. None of them mentioned that it could have induced a recall effect. We have highlighted this point for attention in the instructions for future studies (see [box 1](#)).

Several studies measured GFs with other definitions than the one we used here. For instance, Turnbull *et al* used in their questionnaire 'my GF is 'something is wrong': yes or no, whereas GFs were explained in the instruction booklet as 'GF that the child's illness may be more serious than is superficially apparent'.⁶ Several other studies measuring GFs do miss a detailed and accurate definition of the sense of alarm.^{4,5} In a study regarding the recognition of sepsis in primary care,⁵ the authors did not give details of the concept or definition to which they were referring when using the expression 'GF'. In the questionnaire they used, one item was 'How important were the following patient assessment aspects in the decision to refer?' 'A GF' was one of the possible choices. In our study, we measured GFFs more accurately. The concept of GFs in a closed question is not clear enough, and allows for differences in interpretation of different participants, especially within different languages and cultures. The sense of alarm and the sense of reassurance, as they were defined by Stolper *et al*, were considered, after linguistic validation procedures, as a transcultural concept validated in four languages.^{1,8} In a study measuring the predictive value of GFs for serious infections in children,⁴ the GF was defined as 'an intuitive feeling that something was wrong

even if the clinician was unsure why'. The word 'intuitive' could be a source of misunderstanding as it covers several concepts in cognitive science which sometimes overlap because of different levels of abstraction.^{16–18} Using only this term in research into diagnostic reasoning might be a source of confusion for the participating GPs. We wanted to avoid this possible bias by using well-defined descriptions of the intuitive sense of alarm and of reassurance in our questionnaire.

Using such different definitions and measures of GFs in different contexts, it is not possible to compare the prevalence of the sense of alarm in different studies. The use of the GFQ is a uniform way of measuring the sense of alarm when diagnosing patients in primary care and to determine its prevalence.

Implications for practice and future research

To fill in the questionnaire right after a consultation gives GPs or GP trainees, the opportunity to reflect on their decision-making process. They may thus become aware of GFs and how they play a role in their diagnostic reasoning. Experienced GPs were more likely to report having a GF.⁶ The GFQ is a useful tool for eliciting reflection on diagnostic processes between experienced GPs and trainees.

There is some evidence that the more experience a GP has the more accurate his/her GF is related to the diagnosis of cancer.¹⁹ The GFQ can be used to study this relationship further. In the area of education, how both the sense of alarm and the sense of reassurance play a role in decision-making should be addressed as an important non-analytical track of diagnostic reasoning, especially in general practice.² However, insight into the way GFs are used, and the role of experience should be refined through further studies.

With the final version of the GFQ, prospective observational studies in daily practice can be conducted. A study concerning the accuracy of GPs' sense of alarm when confronted with dyspnoea and/or thoracic pain has already been performed.¹⁰ The results of this study will show the diagnostic test properties, such as the sensitivity and specificity, and the positive and negative likelihood ratios of GPs' sense of alarm, when applied to dyspnoea and chest pain. The relationship between GFs and the diagnosis of cancer can be calculated in this way, just as the relationship between GFs and the outcome of referrals or non-referrals to hospital specialists can be gauged. Knowing to what extent the sense of alarm acts on the decision of a GP in the real context of consultations for non-specific symptoms in primary care is the determining factor.

Author affiliations

¹Department of General Practice, EA 7479 SPURBO, Faculté de Médecine et des Sciences de la Santé, Université de Bretagne Occidentale, Brest, France

²Department of Work and Social Psychology, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands

³Department of Public Health, EA 7479 SPURBO, Faculté de Médecine et des Sciences de la Santé, Université de Bretagne Occidentale, Brest, France

⁴Inserm UMR1078, Faculté de Médecine et des Sciences de la Santé, Université de Bretagne Occidentale, Brest, France

⁵Department of Primary and Interdisciplinary Care, Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium

⁶Department of Family Medicine, Faculty of Health, Medicine and Life Sciences, CAPHRI School for Public Health and Primary Care, Maastricht University, Maastricht, The Netherlands

Acknowledgements We are grateful for the help of the 72 participating GPs and of Alex Gillman who provided medical writing services on behalf of the Collège Brestoïste des Généralistes Enseignants.

Contributors MB, MWJvdW, NG, TM, PVR and ECFS conceived the study, and participated in its design and coordination and helped to draft the manuscript. AD performed and interpreted the statistical analyses. All authors read and approved the final manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent Not required.

Ethics approval Comité Ethique de Brest.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

1. Stolper E, Van Royen P, Van de Wiel M, *et al.* Consensus on gut feelings in general practice. *BMC Fam Pract* 2009;10:66.
2. Stolper E, Van de Wiel M, Van Royen P, *et al.* Gut feelings as a third track in general practitioners' diagnostic reasoning. *J Gen Intern Med* 2011;26:197–203.
3. Stolper E, van Bokhoven M, Houben P, *et al.* The diagnostic role of gut feelings in general practice. A focus group study of the concept and its determinants. *BMC Fam Pract* 2009;10:17.
4. Van den Bruel A, Thompson M, Buntinx F, *et al.* Clinicians' gut feeling about serious infections in children: observational study. *BMJ* 2012;345:e6144.
5. Loots FJ, Arpots R, van den Berg R, *et al.* Recognition of sepsis in primary care: a survey among GPs. *BJGP Open* 2017;1:BJGP-2017-0965.
6. Turnbull S, Lucas PJ, Redmond NM, *et al.* What gives rise to clinician gut feeling, its influence on management decisions and its prognostic value for children with RTI in primary care: a prospective cohort study. *BMC Fam Pract* 2018;19:25.
7. Stolper CF, Van de Wiel MW, De Vet HC, *et al.* Family physicians' diagnostic gut feelings are measurable: construct validation of a questionnaire. *BMC Fam Pract* 2013;14:1.
8. Barais M, Hauswaldt J, Hausmann D, *et al.* The linguistic validation of the gut feelings questionnaire in three European languages. *BMC Fam Pract* 2017;18:54.
9. Oliva B, March S, Gadea C, *et al.* Gut feelings in the diagnostic process of Spanish GPs: a focus group study. *BMJ Open* 2016;6:e012847.
10. Barais M, Barraine P, Scouarnec F, *et al.* The accuracy of the general practitioner's sense of alarm when confronted with dyspnoea and/or thoracic pain: protocol for a prospective observational study. *BMJ Open* 2015;5:e006810.
11. van de Wiel M. Examining expertise using interviews and verbal protocols. *Frontline Learning Research* 2017;5:94–122.
12. Le Reste JY, Coppens M, Barais M, *et al.* The transculturality of 'gut feelings'. Results from a French Delphi consensus survey. *Eur J Gen Pract* 2013;19:237–43.
13. Stolper E, van Royen P, Dinant GJ. The 'sense of alarm' ('gut feeling') in clinical practice. A survey among European general practitioners on recognition and expression. *Eur J Gen Pract* 2010;16:72–4.
14. Hummers-Pradier E, Scheidt-Nave C, Martin H, *et al.* Simply no time? Barriers to GPs' participation in primary health care research. *Fam Pract* 2008;25:105–12.
15. Johnston S, Liddy C, Hogg W, *et al.* Barriers and facilitators to recruitment of physicians and practices for primary care health services research at one centre. *BMC Med Res Methodol* 2010;10:109.
16. Kahneman D, Slovic P, Tversky A. *Judgment under uncertainty: heuristics and biases*. New York: Cambridge University Press, 1982.
17. Glöckner A, Witteman C. Beyond dual-process models: a categorisation of processes underlying intuitive judgement and decision making. *Think Reason* 2010;16:1–25.
18. Hogarth RM. Intuition: a challenge for psychological research on decision making. *Psychol Inq* 2010;21:338–53.
19. Donker GA, Wiersma E, van der Hoek L, *et al.* Determinants of general practitioner's cancer-related gut feelings—a prospective cohort study. *BMJ Open* 2016;6:e012511.