The ‘Surprise question’ in heart failure: a prospective cohort study

Valentina Gonzalez-Jaramillo,1,2 Luisa Fernanda Arenas Ochoa,3 Clara Saldarriaga,3,4 Alicia Krikorian,5 John Jairo Vargas,5,6 Nathalia Gonzalez-Jaramillo,1,2 Steffen Eychmüller,7 Maud Maessen1,7

ABSTRACT

Objective The Surprise Question (SQ) is a prognostic screening tool used to identify patients with limited life expectancy. We assessed the SQ’s performance predicting 1-year mortality among patients in ambulatory heart failure (HF) clinics. We determined that the SQ’s performance changes according to sex and other demographic (age) and clinical characteristics, mainly left ventricular ejection fraction (LVEF) and the New York Heart Association (NYHA) functional classifications.

Methods We conducted a prospective cohort study in two HF clinics. To assess the performance of the SQ in predicting 1-year mortality, we calculated the sensitivity, specificity, positive and negative likelihood ratios, and the positive and negative predictive values. To illustrate if the results of the SQ changes the probability that a patient dies within 1 year, we created Fagan’s nomograms. We report the results from the overall sample and for subgroups according to sex, age, LVEF and NYHA functional class.

Results We observed that the SQ showed a sensitivity of 85% identifying ambulatory patients with HF who are in the last year of life. We determined that the SQ’s performance predicting 1-year mortality was similar among women and men. The SQ performed better for patients aged under 70 years, for patients with reduced or mildly reduced ejection fraction, and for patients NYHA class III/IV.

Conclusions We consider the tool an easy and fast first step to identify patients with HF who might benefit from an advance care planning discussion or a referral to palliative care due to limited life expectancy.

INTRODUCTION

People living with heart failure (HF), especially those in advanced stages of the disease, might present with uncontrolled symptoms such as shortness of breath, pain, sleep disorders and fatigue1; they also often suffer from comorbidities, such as depression and anxiety.2 Advance care planning (ACP) has proven to improve the quality of life and patient satisfaction with end-of-life care for patients with HF by promoting their autonomy concerning medical decisions.3 Identifying those with HF who are in the last year of their life is paramount to guide discussions and determine other strategies that are part of ACP, including possible referrals to specialized palliative care (PC).

The Surprise Question (SQ) is a prognostic screening tool used to identify patients with limited life expectancy. The
SQ is a reflective question a physician or other healthcare providers ask themselves about a patient’s prognosis: ‘would I be surprised if this patient dies within the next 12 months?’ An SQ is positive (+SQ) if the healthcare provider’s answer is ‘no, I would not be surprised’. There are also versions of the SQ within the context of 3 or 6 months. However, 1 year is the most common version.

The SQ’s performance has been evaluated for patients with both oncological and non-oncological diseases, including chronic kidney disease and chronic obstructive pulmonary disease. For patients with oncological diseases, the sensitivity of the 1-year SQ has varied from 58% to 85%. For patients with non-oncological diseases, the performance also varies even within the same disease. Despite the heterogeneity among performance measures, two systematic reviews suggest that, in general, the performance of the SQ predicting 1-year mortality is better for patients with oncological disease than for patients with non-oncological disease. Recently, the SQ’s performance was assessed for an inpatient population with HF and for a population with HF in an emergency setting. The results were promising with a sensitivity of 85% for hospitalised patients and 79% for the emergency setting. However, the SQ’s performance in ambulatory settings, such as HF clinics where patients are more stable, is unknown. Furthermore, understanding the SQ’s performance for men and women is important to ensure equity in the delivery of ACP and PC. To our knowledge, the SQ’s performance for patients with HF, stratified according to the patient’s sex, has not yet been reported. Finally, given that HF clinic admission criteria and samples vary, it is important to understand the SQ’s performance according to other demographic and clinical characteristics to increase the generalisability of the results. Performance of the SQ is likely to differ across subgroups due to previous knowledge of the staff answering the question. For example, evidence has shown a trend of increasing mortality rates with increasing New York Heart Association (NYHA) stage, increasing age and decreasing left ventricular ejection fraction (LVEF).

Therefore, our objectives included (1) assessing the performance of the SQ predicting 1-year mortality among patients in ambulatory HF clinics, and (2) assessing whether performance changes according to sex, age, NYHA classification and LVEF category.

METHODS
This study was conducted and reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology guidelines.

Study design and setting
The prospective cohort included 174 ambulatory patients with HF who were recruited from two HF clinics in Medellín, Colombia between November 2017 and November 2018. One-year vital status was determined by consulting the national mortality register in Colombia. Both clinics are part of tertiary care institutions that are referral centres for patients with cardiovascular disease. They offer comprehensive, multidisciplinary care that includes clinical follow-up by HF cardiologists, nursing education and telephone follow-up, cardiac rehabilitation, and a psychoeducational programme for patients and their families.

Participants
Patients were potentially eligible for the study if they were 18 years or older and existing patients at the HF clinic who had at least two prior consultations. Since the first two consultations provide the cardiologist an opportunity to optimise treatment if necessary and to get to know patients under optimal treatment circumstances according to clinical guidelines, we did not enrol newer HF clinic patients. We enrolled consecutive eligible patients in the study. There were no exclusion criteria.

Ethical aspects
We conducted the study in accordance with the ethical guidelines from the Declaration of Helsinki; our study was approved by the research ethics committees of the institutions involved in the study. Informed consent was collected before participants enrolled in the study.

Data sources and measurements
We obtained sociodemographic characteristics (age, sex and marital status) from electronic medical records, along with values of the clinical variables: LVEF, number of hospitalisations in the last year, presence of cardiac implantable devices, NYHA functional class, comorbidities and current medications. Comorbidities included clinical depression, atrial fibrillation, type 2 diabetes mellitus, kidney disease, lung disease, coronary artery disease, obstructive sleep apnoea and hypothyroidism. Current medications included ACE inhibitors, beta-blockers and angiotensin receptor blockers.

When a patient met the eligibility requirements for study inclusion, the treating cardiologist answered the SQ for that patient. For patients that the cardiologist would not be surprised if they died within the next year, we coded as +SQ. For patients that the cardiologist would be surprised if they died within the next year, we coded as a negative SQ (−SQ).

Statistical methods
To describe the study sample, we used mean and SD to summarise continuous variables in case of normal distribution. In cases of skewed distribution, we used median and IQR. We assessed normality using Q–Q plots. We summarised categorical variables as frequencies and percentages.
To assess the SQ’s performance predicting 1-year mortality, we calculated the sensitivity and specificity, as well as the SQ’s positive (+LR) and negative (–LR) likelihood ratios and the positive (PPV) and negative (NPV) predictive values (online supplemental tables 1 and 2). We interpreted the effect of the +LR on the likelihood of dying within 1 year based on the following classifications: no change if +LR = 1; minimal increase if +LR between 1 and 2; small increase if +LR between 2 and 5; moderate increase if +LR between 5 and 10; and substantial increase if +LR > 10.20

To illustrate how the result of the SQ changes the probability that a patient dies within 1 year, we created Fagan’s nomograms.21 We also conducted a univariable regression to assess the relation between a +SQ and 1-year mortality.

We performed subgroup analysis, comparing groups according to sex, age, LVEF and NYHA functional class. We created a categorical variable for median age (70 years) and another for LVEF (reduced LVEF ≤40%; mildly reduced LVEF 41%–49%; and preserved LVEF ≥50%).22

We performed all analyses with STATA V.15 (Stata Corp, College Station, Texas, USA).

RESULTS

Participants

Of the 184 patients who met the inclusion criteria and were potentially eligible, 178 consented to participate in the study. Among these 178 participants, 4 were excluded because their 1-year vital status was unknown (figure 1).

Table 1 shows baseline demographic and clinical characteristics of the study’s 174 participants. The sample had a median age of 70 (58–77), was predominantly male, had reduced LVEF and NYHA class II. The prevalence of a +SQ was 48%. After 1 year, 20 patients had died, giving an overall mortality rate of 12%.

Performance of the SQ predicting 1-year mortality

After 1 year, mortality among those with a +SQ was 21%; mortality among those with a –SQ was 3% (p<0.001). Participants with a +SQ had 7.5 times higher odds of death at 1 year compared with those with a –SQ (OR 7.6, 95% CI 2.1 to 26.9).

The +LR is the probability that patients with +SQ will die within 1 year divided by the probability that patients with +SQ will be alive in 1 year. The +LR of the SQ was 1.98. The +SQ was nearly twice as likely for patients who died within 1 year than it was for patients who were alive in 1 year (figure 2). According to the classification of the +LR’s effect on the likelihood of dying within 1 year, it is a minor increase. The 1-year mortality rate for our study was 12%. With the pretest probability of a patient dying within 1 year at 12% and a +LR of 1.98, the post-test probability of dying within 1 year is 20% (figure 3).

A +SQ increased the probability that an ambulant patient with HF died within 1 year by 8 percentage points. The –LR was 0.26. The –SQ was nearly four times more likely assigned to participants who were alive in 1 year than it was for patients who died within 1 year.

For our sample, the probability that the SQ correctly identified an individual who would die in the course of a year, the sensitivity was 85% (95% CI 69% to 100%). The probability that the SQ would correctly identify an individual who would survive over a year, the specificity was 57% (95% CI 49% to 65%) (figure 2).

![Figure 1](http://spcare.bmj.com/)

Figure 1  Flow chart of the patients included in the study.
higher +LR, the SQ performed better for participants younger than 70 years (figure 2). Regarding LVEF’s classification, sensitivity, NPV and –LR were better for patients with reduced LVEF. Specificity, PPV and +LR were better for patients with mildly reduced LVEF. We observed the SQ’s worst performance for patients with preserved LVEF (figure 2). Among patients classified as NYHA III/IV, the SQ’s sensitivity was perfect (100%). This subgroup also had the best values of PPV and NPV. However, the SQ’s specificity was very low (31%) for patients at NYHA III/IV classification (figure 2).

Based on Fagan’s nomograms, we accounted for the clinical application of the +LR. We observed the biggest probability changes for patients with +SQ dying within 1 year for those aged under 70 years and for patients with mildly reduced LVEF. Having a +SQ increased the probability of dying within 1 year by 12 and 14 percentage points, respectively (online supplementary figures 1–8). According to the +LR’s effect on the likelihood of dying within 1 year, the +LRs showed a small increase in the likelihood among women; those aged under 70 years and patients with mildly reduced LVEF. In the remaining subgroups, +LR showed a minimal increase in the likelihood of dying within 1 year.

**DISCUSSION**

**Key results**

Our primary objective was to assess the SQ’s performance predicting 1-year mortality among patients in ambulatory HF clinics. With a sensitivity of 85%, the SQ is a good tool to screen ambulatory HF clinic patients who might be in the last year of life. The SQ’s performance predicting 1-year mortality was similar among women and men. We also assessed the SQ’s performance according to sex and other demographic (age) and clinical characteristics. The SQ performed better for patients aged under 70 years; for those with reduced or mildly reduced ejection fraction; and for patients at NYHA III/IV classification. For the whole sample and the different subgroups, +LRs showed a minor or a small increase in the likelihood of a patient dying within 1 year, which is not good enough to consider that a patient has a life expectancy limited to 1 year.

**Performance of the SQ by subgroup**

The risk factors for developing HF differ between men and women, as do responses to treatment, symptom burden, and comorbidities due to both biological and cultural factors. Because of this, sex-specific results should be presented in research. Accounting for sex, determining survival rates for patients with HF has been inconclusive. Initially, the Framingham study showed better survival rates after HF diagnosis for women than for men. Later, other studies suggested worse survival rates for women, which was supported...
by women presenting with HF when they are older and have more comorbidities. However, the most recent evidence suggests that age-adjusted mortality is similar between sexes.

Among the subgroups, the SQ’s best performance was for those aged under 70 years. The main difference between its performance for those older and younger than 70 years was specificity. An age over 70 years likely contributes to a +SQ response from cardiologists, leading to an increase in the proportion of false positives; thus, a reduction in specificity.

The best sensitivity was for the group of patients with reduced LVEF. With more evidence of effective therapies to reduce morbidity and mortality, this type of HF is the most studied and best understood. In addition, evidence has shown that for patients with LVEF regardless of age, the lower the LVEF, the higher the mortality. This might explain why the performance of an intuitive prognostic tool is better for this type of HF. The worst performance was among patients with preserved LVEF. This type of HF is not well understood, and there is no evidence of pharmacological therapy that decreases mortality for patients with preserved LVEF.

As for NYHA classifications, all patients classified at NYHA III/IV who died had a +SQ, which led to a sensitivity of 100%. However, for patients at NYHA III/IV functional class who survived 1 year, the majority also had a +SQ. Perhaps due to previous knowledge that mortality increases with increasing NYHA functional class, cardiologists are more likely to assign +SQ to patients with HF in more advanced stages of NYHA, which is similar to what happens with older patients with HF.

Comparison with previous studies

Previously, the SQ’s performance predicting 1-year mortality had been evaluated for patients with HF in emergency and inpatient settings. In the emergency department, the SQ has a sensitivity of 79% and a specificity of 57% when answered by emergency physicians. For hospitalised patients, the SQ has a sensitivity of 85% and a specificity of 59% when answered by cardiologists. We found that the SQ’s sensitivity (85%) and specificity (57%) for outpatient settings are equal to inpatient settings. Since the SQ’s sensitivity and specificity were the same for decompensated (inpatients) and stable patients (outpatients), it suggests that the SQ’s performance predicting 1-year mortality for patients with HF does not vary significantly. However, to compare the SQ’s performance across settings and its interpretation in clinical practice for individual patients, we would have to compare LRs. No studies assessing the SQ’s performance for populations with HF reported LRs. However, since LRs are calculated using sensitivity and specificity, we estimated them using other studies’ reported test sensitivities and specificities. The +LR and –LR were similar across settings and represent minimal increases in the likelihood of dying within 1 year.
For example, expertise and experience of the staff using it, as well as several aspects of the tool that users should be aware of before applying it in clinical practice.

Since determining sensitivity was the main criterion to evaluate the SQ's performance, we consider it is an acceptable prognostic screening tool for patients with HF. However, there remains a significant percentage of patients who will survive longer than 1 year but still would benefit from ACP or PC.

When there is high test sensitivity, fewer false negatives occur, which increases the chance that patients with HF in need of ACP or PC will receive these services. Even in scenarios when the SQ predicts mortality well, using life expectancy as the sole criterion for assessing the need for ACP or PC is limiting. We consider the SQ can be used as a screening tool to initiate ACP or refer to PC for patients with a life expectancy of less than 1 year. However, we also consider the parallel use of needs assessment tools for patients with life expectancy of more than 1 year. For example, two recent systematic reviews of available tools to assess PC needs in patients with HF concluded that the Needs Assessment Tool: Progressive Disease-Heart Failure (NAT: PD-HF) was the most appropriate tool to determine the unmet needs of patients with HF.

The NAT: PD-HF offers an alternative solution to several previously discussed points: it is not based on survival prognosis or severity factors, but it comprehensively evaluates different spheres.

The NAT: PD-HF is made of questions to determine a patient’s physical and psychological symptoms, daily life activity limitations, spiritual concerns, financial or legal concerns, and health-related information needs. Although the NAT: PD-HF does not focus on PC needs or referrals for specialised PC services, it does assess patients’ unmet needs and matches those needs with appropriate services, including specialised PC and other services. Finally, the NAT: PD-HF assesses the patients’ and the caregivers’ needs, including the caregiver’s ability to take care of the patient. However, in clinical situations where there is not enough time to gather answers to the NAT: PD-HF’s comprehensive question sets, there is enough time for the clinician to ask themselves the singular SQ, which is better than no needs screening at all.

Interpretations for clinical practice
When there is high test sensitivity, fewer false negatives occur, which increases the chance that patients with HF in need of ACP or PC will receive these services. Since determining sensitivity was the main criterion to evaluate the SQ’s performance, we consider it is an acceptable prognostic screening tool for patients with HF. Yet, there remains a significant percentage of patients with HF within the last year of their life who might be left out (15% of the whole sample) and might benefit from ACP or PC. Both in this population and in the literature, the SQ predicts 1-year mortality well among patients with HF. However, there are several aspects of the tool that users should be aware of before applying it in clinical practice.

First, the SQ’s performance depends on the clinical expertise and experience of the staff using it, as well as their knowledge of the patient with HF. For example, Straw et al conducted a study that evaluated the SQ’s performance among an inpatient population with HF. They showed that the SQ’s sensitivity decreased from 85% when cardiologists used the tool to 75% when physicians in training used it; and from 90% when HF nurses used the SQ to 66% among non-specialist nurses. Second, screening strategies such as the SQ do not afford assessment of the complexity of patients’ needs or the level of training that professionals should have to address ACP discussions or provide PC.

Finally, a major limitation of this screening tool is that it potentially excludes patients who will survive longer than 1 year but still would benefit from ACP or PC. Even in scenarios when the SQ predicts mortality well, using life expectancy as the sole criterion for assessing the need for ACP or PC is limiting. We consider the SQ can be used as a screening tool to initiate ACP or refer to PC for patients with a life expectancy of less than 1 year. However, we also consider the parallel use of needs assessment tools for patients with life expectancy of more than 1 year. For example, two recent systematic reviews of available tools to assess PC needs in patients with HF concluded that the Needs Assessment Tool: Progressive Disease-Heart Failure (NAT: PD-HF) was the most appropriate tool to determine the unmet needs of patients with HF.

The NAT: PD-HF offers an alternative solution to several previously discussed points: it is not based on survival prognosis or severity factors, but it comprehensively evaluates different spheres.

The NAT: PD-HF is made of questions to determine a patient’s physical and psychological symptoms, daily life activity limitations, spiritual concerns, financial or legal concerns, and health-related information needs. Although the NAT: PD-HF does not focus on PC needs or referrals for specialised PC services, it does assess patients’ unmet needs and matches those needs with appropriate services, including specialised PC and other services. Finally, the NAT: PD-HF assesses the patients’ and the caregivers’ needs, including the caregiver’s ability to take care of the patient. However, in clinical situations where there is not enough time to gather answers to the NAT: PD-HF’s comprehensive question sets, there is enough time for the clinician to ask themselves the singular SQ, which is better than no needs screening at all.

Strengths and limitations of this study
The low proportion of patients classified as NYHA III/IV might be a limit of the generalisability of our results. The risk of mortality increases with a higher NYHA classification. Among our population, 80% of the patients were classified as NYHA I or II. As expected, the mortality rate was low compared with what has been reported in other HF clinics where mortality is around 30%. However, as we conducted different subgroup analyses, including analysis according to NYHA functional class, we provide different analyses.
that can be assessed according to each HF clinic population. As most encounters between healthcare personnel and patients occur in ambulatory settings, a strength of this study is the contribution of the SQ’s psychometric profile for ambulatory patients with HF. In addition, when screening for limited life expectancy and the need to initiate ACP, since the outpatient population are more stable patients, ACP needs may be overlooked. The systematic use of a tool such as the SQ could help identify patients with HF eligible for end-of-life care without being limited by having ACP discussion or making decisions within unstable medical contexts. Furthermore, to support better generalisation, this study provides substantial data regarding subcategories.

CONCLUSION

The SQ showed a good sensitivity predicting 1-year mortality for patients in ambulatory HF clinics. However, the likelihood of dying within 1 year increases little when having a +SQ. We suggest that the SQ can be used as a starting point to identify patients who might benefit from having an ACP discussion or a referral to PC due to limited life expectancy toward a patient’s end of life.

Acknowledgements We thank Christopher Owen Ritter and Kristin Bivens, scientific editors for the Institute of Social and Preventive Medicine (ISPM) at the University of Bern, for their editorial contributions to this article. We also thank Lukas Bütikofer, senior statistician at ISPM, for his input regarding the study’s statistical analysis.

Contributors LF AO designed the study and collected the data. VG-J performed the analysis, drafted the manuscript and designed the figures. NG-J designed the figures and contributed to the analysis of results. LF AO, CS, AK, JJV, NG-J, SE and MM did a critical revision of the manuscript. All authors discussed the results and commented on the manuscript. VG-J and MM are responsible for the overall content of the manuscript.

Funding This project received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement no 801076, through the SSPH+Global PhD Fellowship Programme in Public Health Sciences (GlobalP3HS) of the Swiss School of Public Health.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Data are available upon request to the corresponding author.

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/.

ORCID iDs
Valentina Gonzalez-Jaramillo http://orcid.org/0000-0002-2470-7022
Alicia Krikorian http://orcid.org/0000-0003-2118-5692

REFERENCES

13 Aaronson EL, George N, Ouchi K, et al. The surprise question can be used to identify heart failure patients in the emergency department who would benefit from palliative care. J Pain Symptom Manage 2019;57:944–51.
29 Ponikowski P, Voors AA, Anker SD, et al. 2016 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure: the task force for the diagnosis and treatment of acute and chronic heart failure of the European Society of cardiology (ESC), developed with the special contribution of the heart failure association (HFA) of the ESC. Eur J Heart Fail 2016;18:891–975.