Accepted Manuscript

British Journal of General Practice

Online experiment comparing GPs' antibiotic prescribing decisions to a clinical prediction rule

Nurek, Martine; Hay, Alastair; Kostopoulou, Olga

DOI: https://doi.org/10.3399/BJGP.2020.0802

To access the most recent version of this article, please click the DOI URL in the line above.

Received 27 August 2020
Revised 25 August 2022
Accepted 02 September 2022

TITLE

Online experiment comparing GPs' antibiotic prescribing decisions to a clinical prediction rule

AUTHORS

Martine Nurek, PhD (0000-0002-4252-4692)*,[1], Alastair D Hay, FRCGP (0000-0003-3012-375X)[2], and Olga Kostopoulou, PhD (0000-0001-9643-0838)[1]

[1] Imperial College London, Department of Surgery and Cancer, Faculty of Medicine, 5th floor Medical School Building, St Mary's Campus, Norfolk Place, London W2 1PG, UK.

[2] University of Bristol, Centre for Academic Primary Care, Bristol Medical School: Population Health Sciences, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS.

* Correspondence concerning this article should be addressed to Martine Nurek, Imperial College London, Department of Surgery and Cancer, 5th floor Medical School Building, St Mary's Campus, Norfolk Place, London W2 1PG, UK (email: m.nurek@imperial.ac.uk; phone: +44 (0)20 759 43062)

ABSTRACT

Background: The"STARWAVe" clinical prediction rule (CPR) uses seven factors to guide risk assessment and antibiotic prescribing in children with cough (**S**hort illness duration, **T**emperature, **A**ge, **R**ecession, **W**heeze, **A**sthma, **V**omiting).

Aim: To assess the influence of STARWAVe factors on General Practitioners' (GPs) unaided risk assessments and prescribing decisions. We also explored two methods of obtaining risk assessments and tested the impact of parental concern.

Design and setting: Experiment comprising clinical vignettes administered to 188 UK GPs online.

Method: GPs were randomly assigned to view 32 (of 64) vignettes depicting children with cough. Vignettes varied the STARWAVe factors systematically. Per vignette, GPs assessed risk of deterioration in one of two ways (sliding scale vs. risk category selection) and indicated whether they would prescribe antibiotics. Finally, they saw an additional vignette, suggesting that the parent was concerned. Using mixed-effects regressions, we measured the influence of STARWAVe factors, risk elicitation method, and parental concern on GPs' assessments and decisions.

Results: Six STARWAVe risk factors correctly increased GPs' risk assessments (bs$_{\text{sliding-scale}} \geq 0.66$, ORs$_{\text{category-selection}} \geq 1.61$, ps≤ 0.001) while one incorrectly reduced them (short duration: $b_{\text{sliding-scale}} = -0.31$, $\text{OR}_{\text{category-selection}} = 0.75$, ps≤ 0.039). Conversely, one STARWAVe factor increased prescribing odds (fever: OR=5.22, $p<0.001$) while the rest either reduced them (short duration, age, recession: ORs≤ 0.70, ps<0.001) or had no significant impact (wheeze, asthma, vomiting: ps≥ 0.065). Parental concern increased risk assessments ($b_{\text{sliding-scale}} = 1.29$, $\text{OR}_{\text{category-selection}} = 2.82$, ps≤ 0.003) but not prescribing ($p=0.378$).

Conclusion: GPs use some, but not all, STARWAVe factors when making unaided risk assessments and prescribing decisions. Such discrepancies must be considered when introducing CPRs to clinical practice.

Keywords:
STARWAVe, antimicrobial stewardship, decision support, primary care, general practice

HOW THIS FITS IN

The STARWAVe clinical prediction rule is a promising antimicrobial stewardship tool that could dramatically decrease unnecessary prescribing in children with RTIs. Our findings suggest, however, that GPs do not spontaneously interpret patient characteristics, symptoms, and signs in accordance with STARWAVe – particularly in their prescribing decisions. Discordance between GPs and STARWAVe must be addressed if this tool is to bring about meaningful change in practice.

INTRODUCTION

Antimicrobial resistance (AMR) is a major threat to public health, with AMR-infections claiming an estimated 700,000 lives per year globally.[1] This figure is expected to increase to 10 million by 2050 if no action is taken.[1] Central to the UK five-year action plan to tackle AMR is prudent antibiotic prescribing.[1] Most NHS prescriptions are issued in primary care,[2] often for childhood respiratory tract infections (RTIs).[3,4] It is known that antibiotics offer limited benefits in such cases,[5-7] but children are perceived as vulnerable and the risk of future deterioration leads to "defensive" antibiotic use.[4,8-10]

Risk of future deterioration in children with RTIs can now be estimated with reasonable accuracy, using the validated "STARWAVe" clinical prediction rule (CPR).[4] STARWAVe uses seven factors – **S**hort illness duration (≤3 days), **T**emperature (≥37.8℃), **A**ge (<2 years), **R**ecession, **W**heeze, **A**sthma, **V**omiting – to differentiate children at "very low" (0.3%, ≤1 factor present), "normal" (1.5%, 2-3 present) and "high" (11.8%, ≥4 present) risk of hospitalisation within a month.[4] Consistent with national guidelines,[11] STARWAVe supports antimicrobial treatment only in high risk cases. With only 3% of acute childhood RTIs falling into the high risk category,[4] widespread STARWAVe-uptake could dramatically decrease unnecessary prescribing.

One way to increase uptake is to integrate STARWAVe into patients' electronic health records. Indeed, within-consultation STARWAVe decision support has been incorporated into a complex behavioural intervention (currently at clinical trial) intended to improve management of childhood RTIs.[12] It is known however that risk algorithms can be met with distrust, especially if they conflict with the practitioner's own clinical judgment.[13] It is therefore important to understand whether and how practitioners deviate from STARWAVe, and how they interpret its clinical factors.

Research questions and hypotheses

We investigated whether and how GPs' risk assessments and prescribing decisions differ from those suggested by STARWAVe. As this was an extension and

improvement of a previous study[14] (for more information, see our approved protocol), we expected to replicate our previous findings; specifically:

- Vomiting and wheeze would both be associated with higher risk assessments and antibiotic prescribing odds;
- Younger patient age would be associated with higher risk assessments but would not influence prescribing;
- Shorter duration would be associated with lower risk assessments and prescribing odds.

We had no hypotheses concerning the remaining three STARWAVe factors, which were not investigated in the previous study.

We also addressed three secondary research questions:

1) *Which of two risk elicitation methods will maximise alignment between GPs' risk assessments and prescribing decisions?* We had no hypotheses in this regard.

2) *Which of two risk elicitation methods will maximise alignment between GPs and STARWAVe?* We had no hypotheses in this regard.

3) *How might parental concern influence risk assessments and prescribing decisions?* We expected prescribing to increase when parental concern was present (vs. absent) from a clinical vignette.[15] We had no hypotheses as to whether/how parental concern might influence risk assessments.

METHOD

Design and materials

Materials were 64 vignettes depicting children with cough (Appendix 1). Each child was described in terms of the STARWAVe factors, which were manipulated in a half-fractional factorial design (2^{7-1}; Table 1). For face validity, patient sex was included. For external validity, vignettes were based on real patients (further details in protocol).

We divided the vignettes into two sets (set A and B) and subdivided each set into two surveys. GPs were randomly assigned to either set A or B and saw the two surveys 24hrs apart (order counterbalanced across GPs). The number of very low, normal, and high-risk cases was consistent across sets and surveys (Appendix 2).

We also constructed seven "parental concern" vignettes, by adding the phrase "the parent is quite concerned" to each of the seven very low risk cases (Appendix 3). Each GP saw one parental concern vignette, selected at random from the very low risk cases in his/her *unseen* set. This vignette was always presented last, so as not to influence responses to the "primary" vignettes.

Procedure

GPs received an invitation e-mail, with a link to an Expression of Interest form (EOI, Appendix 4). Those who expressed interest were e-mailed a link to the study website. After providing consent and reading an introduction to the study (Appendix 5), GPs saw 16 vignettes in a random order. Per vignette, they were asked three questions:

1) *In your opinion, what is the probability that this child would deteriorate, requiring hospital admission?*
   Half of the sample (randomly selected) expressed their answer as a percentage, using a sliding scale (Figure 1A). The scale was capped at 20% because the probability rarely exceeds 17% in this cohort.[4] The rest selected between three risk categories ("extremely low, around 0.3%", "low, around 1.5%", "moderate or high, around 7% and above"; Figure 1B). These categories correspond to STARWAVe's three levels of risk, but the STARWAVe labels ("very low", "normal", "high"), were replaced by GP-appointed labels, elicited in a pilot study (Appendix 6). For consistency, we adopt the terminology of *level 1* (lowest), *level 2* (middle), and *level 3* (highest) risk, whether referring to GPs' responses or STARWAVe categories.

2) *In your clinical judgment, what would be the best course of action?*

We provided three options for patient management: “prescribe antibiotics”, “arrange to GP review within 24hrs” and “admit for paediatric assessment”. GPs could select all that applied (or none). When “prescribe antibiotics” was selected, GPs were asked whether the antibiotic would be “immediate” or “delayed”.

3) *Please enter any additional comments (optional).*

Twenty-four hours later, GPs were e-mailed a link to a second survey, comprising 16 previously-unseen vignettes and one parental concern vignette. Response scales were identical to those seen in the first survey. The procedure is presented graphically in Appendix 7.

Surveys were hosted online using Qualtrics. After completing both surveys, GPs received £50 via bank transfer and the NIHR-CRN (National Institute for Health Research Clinical Research Network) gave £50 to each GP’s practice. Data were collected Apr-Oct 2021. Departures from the approved protocol are reported in Appendix 8.

Analysis

Primary research question

To investigate the effect of the STARWAVe factors on risk assessments and prescribing decisions, three mixed-effects regression models were built, with random intercept and slope by participant:

1) Continuous risk assessments (cast on a sliding scale) were regressed upon patient age (continuous), duration (continuous), vomiting, asthma, recession, fever, and wheeze (0=absent, 1=present; linear model).

2) Risk category selections (0=*level 1*, 1=*level 2*, 2=*level 3*) were regressed upon the same (ordinal logistic model).

3) Prescribing decisions were dichotomised (0=no/delayed prescription, 1=immediate prescription), in line with national guidelines (which treat "no prescription" and "delayed prescription" interchangeably),[11] and regressed upon the same factors (binary logistic model). For exploratory purposes, the model was repeated with a 3-category ordinal dependent variable (0=no prescription, 1=delayed prescription, 2=immediate prescription; ordinal logistic model).

All models were repeated with age and duration treated as binary, in keeping with the STARWAVe CPR (age=1 if <2 years; duration=1 if $\leq$ 3 days; otherwise 0).

Secondary research questions

To measure the alignment between GPs' risk assessments and prescribing decisions – and compare this across risk response modes – we classified immediate prescriptions as either "consistent" or "inconsistent" with the GP's risk assessment. As prescribing was not the only management option available to GPs, we also classified each case as "appropriately" or "inappropriately" managed, relative to the GP's risk assessment. (For full details of these classification systems, see protocol). These two variables were then regressed upon risk response mode, using mixed-effects random-intercept binary logistic regression.

To measure GPs' alignment with STARWAVe – and compare this across risk response modes – risk assessments, immediate prescriptions, and broader management decisions were each classified as either "consistent" or "inconsistent" with STARWAVe (full details in protocol), and regressed upon response mode using mixed-effects, random-intercept binary logistic regression.

To assess the impact of parental concern, we selected only STARWAVe-identified *level 1* cases (with and without parental concern) and built three mixed-effects, random-intercept regression models. First, risk assessments were regressed upon parental concern (0=absent, 1=present), separately for those who used a sliding scale (linear model) vs. category selection (ordinal logistic model). Thereafter, all prescribing decisions (0=no/delayed prescription, 1=immediate prescription) were regressed upon

parental concern (binary logistic model). Parental concern vignettes were not included in any other analyses.

Analyses were performed using Stata/MP 17.0. Mixed-effects regressions were conducted using the "mixed" (linear), "melogit" (binary), and "meologit" (ordinal) commands.[16] Violations of the proportional odds assumption were addressed using the "gologit2" command.[17,18]

Participants

GPs practising in the United Kingdom were eligible to participate, as were ST3/4 GP trainees. Participants were recruited via the NIHR-CRN, who circulated our invitation e-mail to practices across England.

Power analyses were conducted for each hypothesis (Appendix 9). These suggested that 429 [88] GPs were needed to assess the effect of the STARWAVe factors on categorical [continuous] risk assessments. As 429 GPs was not feasible (GPs being a difficult-to-reach population), we deemed it practical to recruit two groups of 88 GPs (one per risk response scale; *N*=176 in total). Assuming a dropout rate of 2%, we aimed to recruit an additional four, yielding *n*=90 per risk response scale.

RESULTS

306 GPs completed the EOI. This is higher than our intended sample size because – in our experience – those who express interest do not always go on to complete the study. Indeed, only 199 completed the first survey (65%); of these, 188 went on to complete the second survey (95%). The 11 GPs who did not complete the second survey were excluded from the analysis. Sample characteristics appear in Table 2.

Descriptive statistics

On average – excluding parental concern vignettes – the "sliding-scale group" estimated risk of hospitalisation to be 4.1% (*SD*=3.92), 7.9% (*SD*=5.32), and 11.9%

(*SD*=5.50) for cases classified by STARWAVe as *level 1, level 2,* and *level 3*, respectively. Figure 2 compares their risk assessments (categorised as *level 1* if <1%, *level 2* if 1%-6.9%, or *level 3* if $\geq$ 7%) to those of the "category-selection group" and STARWAVe. It suggests that the sliding-scale group tended to overestimate risk, frequently assigning a risk level of 2 to cases that STARWAVe deemed *level 1* (66%, 217/329), and a risk level of 3 to cases that STARWAVe deemed *level 2* (51%, 835/1645). The category-selection group appeared less likely to overestimate risk, with most GPs selecting *level 1* for STARWAVe-identified *level 1* cases (62%, 205/329) and *level 2* for STARWAVe-identified *level 2* cases (51%, 831/1645).

GPs' management selections are shown in Figure 3. In STARWAVe-identified *level 1* cases, the commonest course of action was to review the patient (selected 51%/64% of the time in the sliding-scale/category-selection group), though many preferred to take no action at all (35%/24%). In STARWAVe-identified *level 2* cases, the commonest action was again GP review (55%/63%), though a substantial minority opted to admit the patient for paediatric assessment (30%/30%). Admission for paediatric assessment was the most-selected option for STARWAVe-identified *level 3* cases (64%/66%), though GP review remained prominent (39%/38%). Across the board, antibiotics were infrequently prescribed and tended to be immediate (15%/12% of all cases) rather than delayed (4%/4%).

Primary research question

The effect of STARWAVe risk factors on GPs' risk assessments is shown in Table 3 (columns 1 and 2). In keeping with STARWAVe (green cells), presence of fever, vomiting, asthma, recession, and wheeze significantly increased risk assessments, as did younger age (i.e., older age reduced them). Inconsistent with STARWAVe (red cells), longer duration significantly increased risk assessments.

Table 3 also shows the effect of the STARWAVe factors on prescribing decisions. Consistent with STARWAVe (green cell), fever significantly increased prescribing odds; inconsistent with STARWAVe (red cells), longer duration and older age

increased them, while recession reduced them. Vomiting, asthma, and wheeze had no significant impact (orange cells).

Secondary research questions

Alignment between GPs' risk assessments and prescribing decisions

Half of GPs' prescriptions were inconsistent with their own risk assessments; i.e., administered in cases that they perceived to be *level 1* or *level 2* (51%, 407/797). This occurred significantly more often in the category-selection group (64% of prescriptions, 223/348) than the sliding-scale group (41%, 184/449; OR=4.25 [2.16-8.39], p<0.001). More broadly, management selections were inconsistent with subjective risk assessments 42% of the time (2549/6016), again more often in the category-selection group (50%, 1513/3008) than the sliding-scale group (34%, 1036/3008; OR=2.11 [1.65-2.69], p<0.001; Appendix 15).

Alignment between GPs and STARWAVe

Risk assessment was inconsistent with STARWAVe 44% of the time (2653/6016), though this was less frequent in category-selection group (42% inconsistent, 1254/3008) than the sliding-scale group (47%, 1399/3008; OR=0.82 [0.73-0.92], p=0.001). As Figure 2 suggested, the sliding-scale group tended to overestimate risk (37% of cases, 1110/3008) rather than underestimate it (10%, 289/3008), while the category-selection group did both (22% and 20%, 666/3008 and 588/3008).

While risk assessment was better-aligned with STARWAVe in the category-selection group, prescribing decisions were not. Most immediate prescriptions (69%, 550/797) were unnecessary according to STARWAVe, and this proportion did not differ by risk response mode (72% (250/348) vs. 67% (300/449); OR=1.28 [0.92-1.77], p=0.145). In fact, the category-selection group was more likely than their counterparts to manage the patient inappropriately, according to STARWAVe (59% vs. 55% of cases (1774/3008 vs. 1656/3008); OR=1.17 [1.05-1.32], p=0.006; Figure 3).

Parental concern

Adding parental concern to STARWAVe-identified *level 1* vignettes significantly increased risk assessments in both groups ($b_{\text{sliding-scale}}$=1.29 [0.44-2.14], p=0.003; $OR_{\text{category-selection}}$=2.82 [1.77-4.49], $p \leq 0.001$; proportional odds assumption met with $\chi^2(1)=0.01$, p=0.938). In the sliding-scale group, the mean risk assessments were 5.33% (concern present) vs. 4.1% (absent); in the category-selection group, the median/modal risk assessments were *level 2* (concern present) vs. *level 1* (absent).

Parental concern did not influence prescribing (OR=0.74 [0.39-1.44], p=0.378), but did increase the odds of GP review (OR=2.88 [1.90-4.36], p<0.001) and paediatric assessment (OR=2.82 [1.72-4.63], p<0.001). When parental concern was present [absent] in these vignettes, GPs prescribed 7% [9%] of the time, reviewed the patient 76% [57%] of the time, and admitted the patient 17% [7%] of the time (Appendix 16).

DISCUSSION

Summary and comparison with literature

We compared GPs' risk assessments and antibiotic prescribing decisions to the STARWAVe CPR. Risk assessments were aligned with STARWAVe roughly half of the time, though degree of alignment varied by risk response mode (category-selection 58% vs. sliding-scale 53%). Risk response mode also influenced GPs' deviations from STARWAVe: the sliding-scale group tended to overestimate (vs. underestimate) risk, while the category-selection group did both. In a previous study,[14] we employed a third risk response mode and observed yet a third type of deviation (underestimation of risk), suggesting that risk elicitation mode can influence risk perceptions ("reactivity"[19,20]).

While GPs frequently perceived risk to be high (*level 3*), the prescribing rate was low (13%), with GPs preferring further assessment (review/referral) to immediate prescribing. Still, many prescriptions were unnecessary relative to GPs' own risk assessments (51%) and STARWAVe's (69%). These findings were consistent across risk response modes and consistent with our previous study, where prescriptions

(administered 15% of the time) were usually unnecessary relative to GPs' (78%) and STARWAVe's (83%) risk assessments.

This suggests a disconnect between risk assessments and prescribing decisions, which was also apparent in GPs' interpretation of STARWAVe's risk factors. Younger patient age and recession increased GPs' risk assessments (in keeping with STARWAVe) but reduced prescribing odds. Vomiting, asthma, and wheeze likewise increased risk assessments (in keeping with STARWAVe) but did not affect prescribing. Only long illness duration and fever increased both risk perceptions and prescribing, but the former runs counter to STARWAVe (which posits short illness duration as a risk factor).

Notably, the effect of the STARWAVe factors on risk assessments was remarkably consistent across response modes (sliding-scale vs. category-selection). It was also consistent with our previous study[14] and STARWAVe itself (with only one factor – illness duration – incorrectly informing GPs' risk assessment). In contrast, the effect of the STARWAVe factors on prescribing decisions was not consistent with our previous study[14] or STARWAVe (with only one factor – fever – correctly informing GPs' decisions). GPs' interpretation of the STARWAVe factors would thus appear to be stable and largely appropriate when it comes to risk assessment, but variable and largely inappropriate when it comes to prescribing decisions.

In very low risk cases (*level 1*), parental concern raised GPs' risk assessments but not prescribing odds, speaking again to the disconnect between risk assessment and prescribing. The non-association between parental concern and prescribing was somewhat surprising, given that parental pressure for antibiotics is known to increase their likelihood.[9,15,21-23] Parental concern, however, is not specific to antibiotics, and it did propel GPs to take other types of action, including 24hr review and admission for paediatric assessment (both unnecessary in very low risk cases cases).

Implications for research/practice

If STARWAVe is to be provided as a decision aid to GPs, then several issues require address. Firstly, risk of hospitalisation guides prescribing in the STARWAVe model; our findings suggest that this may not be true for GPs. Consequently, a STARWAVe-based decision aid may fail to bring about meaningful change. As mentioned in the *Introduction*, a multi-component intervention that includes STARWAVe decision support is currently at clinical trial,[12] but its success/failure cannot confidently be attributed to a single component. Therefore, further work is needed to isolate the effect of STARWAVe on prescribing practices.

Secondly, decision aid developers will need to consider how best to present risk estimates to GPs. Our findings suggest that risk elicitation format can determine explicit risk assessments. Risk assessments cast on a sliding scale were inflated relative to those cast via category selection, but they were also better-aligned with GPs' own management decisions (suggesting a certain "fidelity" or "construct validity") and STARWAVe's recommended management pathways (suggesting potential normative value). Presenting risk scores as point-estimates on a continuous scale may thus be more effective – and more likely to influence decisions – than presenting them as categories (e.g., "very low", "high").

Thirdly, we find consistent evidence to suggest that GPs' interpretation of illness duration runs counter to STARWAVe. Qualitative work investigating why GPs prescribe antibiotics for longer (vs. shorter) illnesses could return valuable insights. GPs may be adopting a "wait-and-see" approach to prescribing, in an attempt to reduce prescriptions.[14] Alternatively, they may be concerned that a prolonged infection has/will become bacterial, even if it began as viral. Either way, decision aid developers need to be aware that this risk factor is counter-intuitive to GPs and may require explanation.

Strengths and limitations

Methodological rigor was the guiding principle of this study, which improved upon a previous one[14] by 1) basing patient vignettes upon real patients, 2) varying all seven STARWAVe factors, 3) testing the effects of two new risk elicitation modes, 4)

ensuring a comprehensive and clinically plausible set of options for patient management, and 5) testing the effect of a non-clinical factor (parental concern). More broadly, we used a robust experimental approach that is under-represented in this area of research and predefined our hypotheses, procedure, and analyses in an approved protocol.

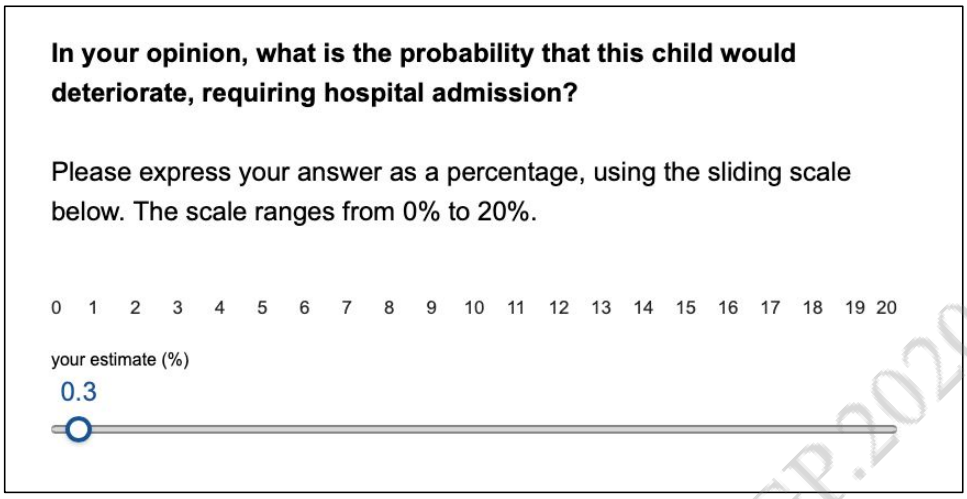
Despite our attempts to improve ecological validity, patients appeared 'on paper' rather than in person. We therefore appreciate that some of our findings may fail to translate to clinical practice, where situational variables such as time pressure are known to play a role.[8,9,22,24-26] Presently, for example, GPs elected to review (within 24hrs) roughly half the patients seen; whether they would or could review these patients when faced with the pressures of routine practice is questionable.

We are also aware that analyses limited to the category-selection group (Table 3, column 2) were underpowered and therefore exploratory. While findings were comparable to those of the sliding-scale group, they would benefit from replication in a larger study. Future work might also assess whether present findings extend to other primary care clinicians who manage childhood RTIs, including nurses, paramedics, and pharmacists.

Despite these limitations, the study highlights the importance of examining how practitioners appraise information and assess risk when designing decision support tools. When discrepancies are observed, such tools should be introduced carefully and explained to practitioners, so that they do not result in loss of trust and limited use.

FIGURE 1

A)



B)

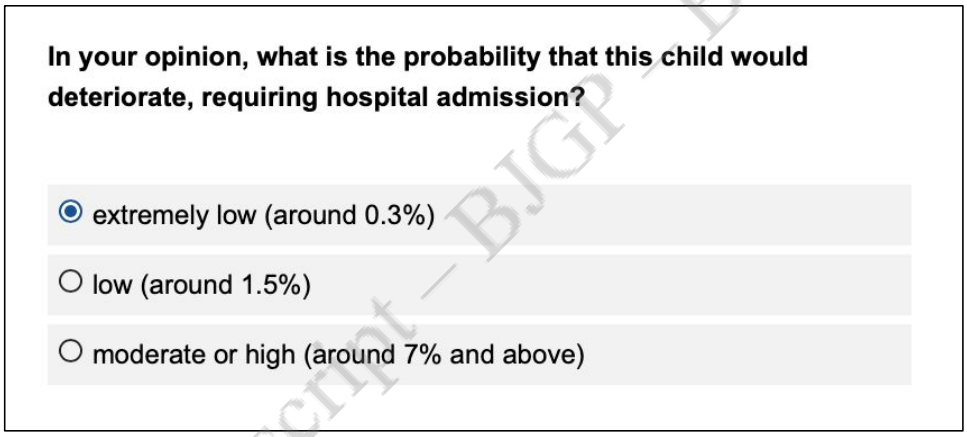


Figure 1. Response scales used to elicit perceived risk of hospitalisation. Mock responses are provided for illustrative purposes: "0.3%" on the sliding scale (A) and "extremely low, around 0.3%" on the category selection scale (B).

FIGURE 2

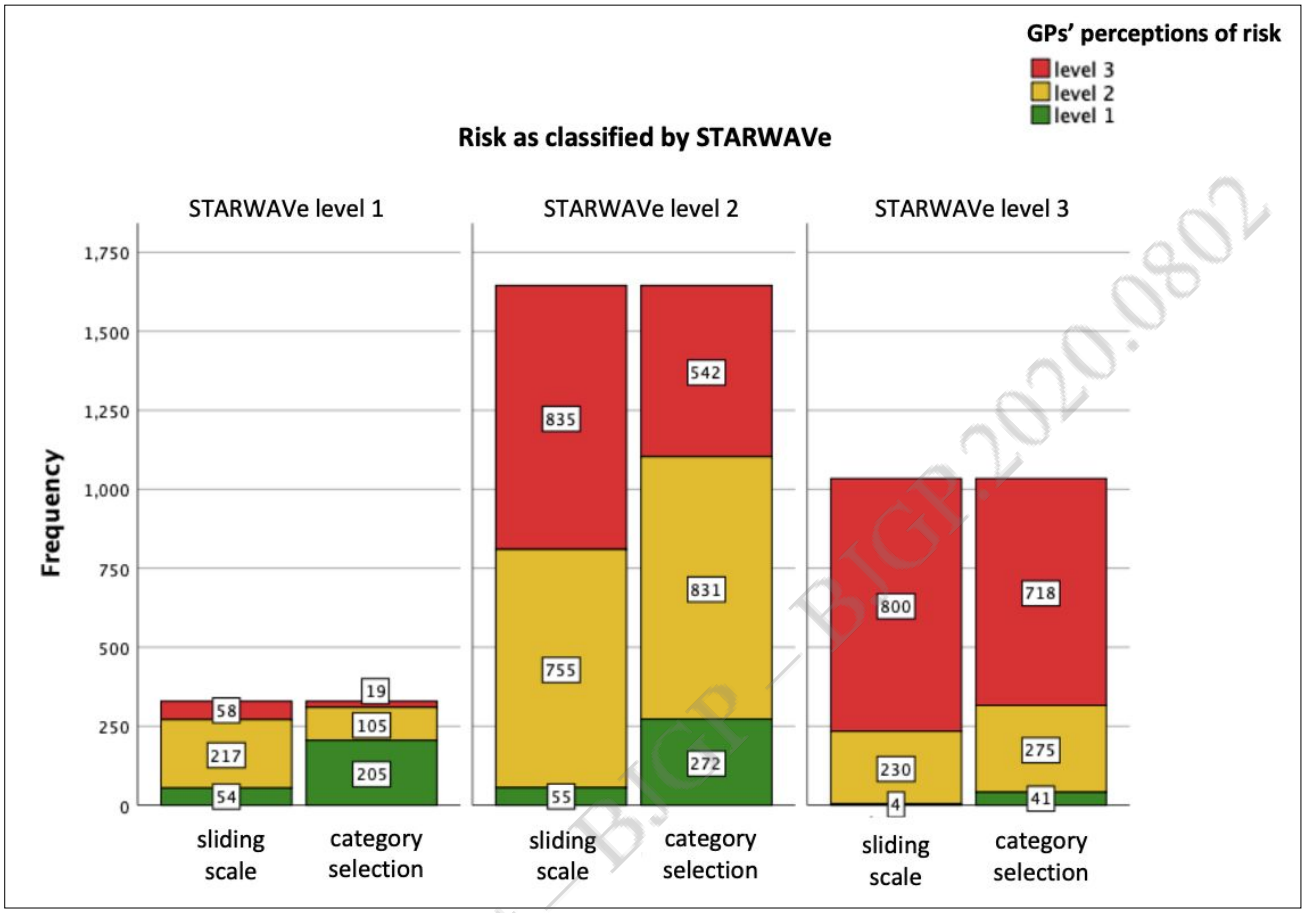


Figure 2. GPs' risk assignations (level 1 vs. level 2 vs. level 3), by STARWAVe risk classification and risk response mode. Risk assessments cast on a 0-20% sliding scale were classified as *level 1* if <1%, *level 2* if 1%-6.9%, or *level 3* if ≥7%. Risk assessments cast via category selection were classified as *level 1* if "extremely low, around 0.3%" was selected, *level 2* if "low, around 1.5%" was selected, or *level 3* if "moderate or high, around 7% and above" was selected. The number of cases that STARWAVe classified as *level 1, level 2,* and *level 3* was 329, 1645, and 1034, respectively.

FIGURE 3

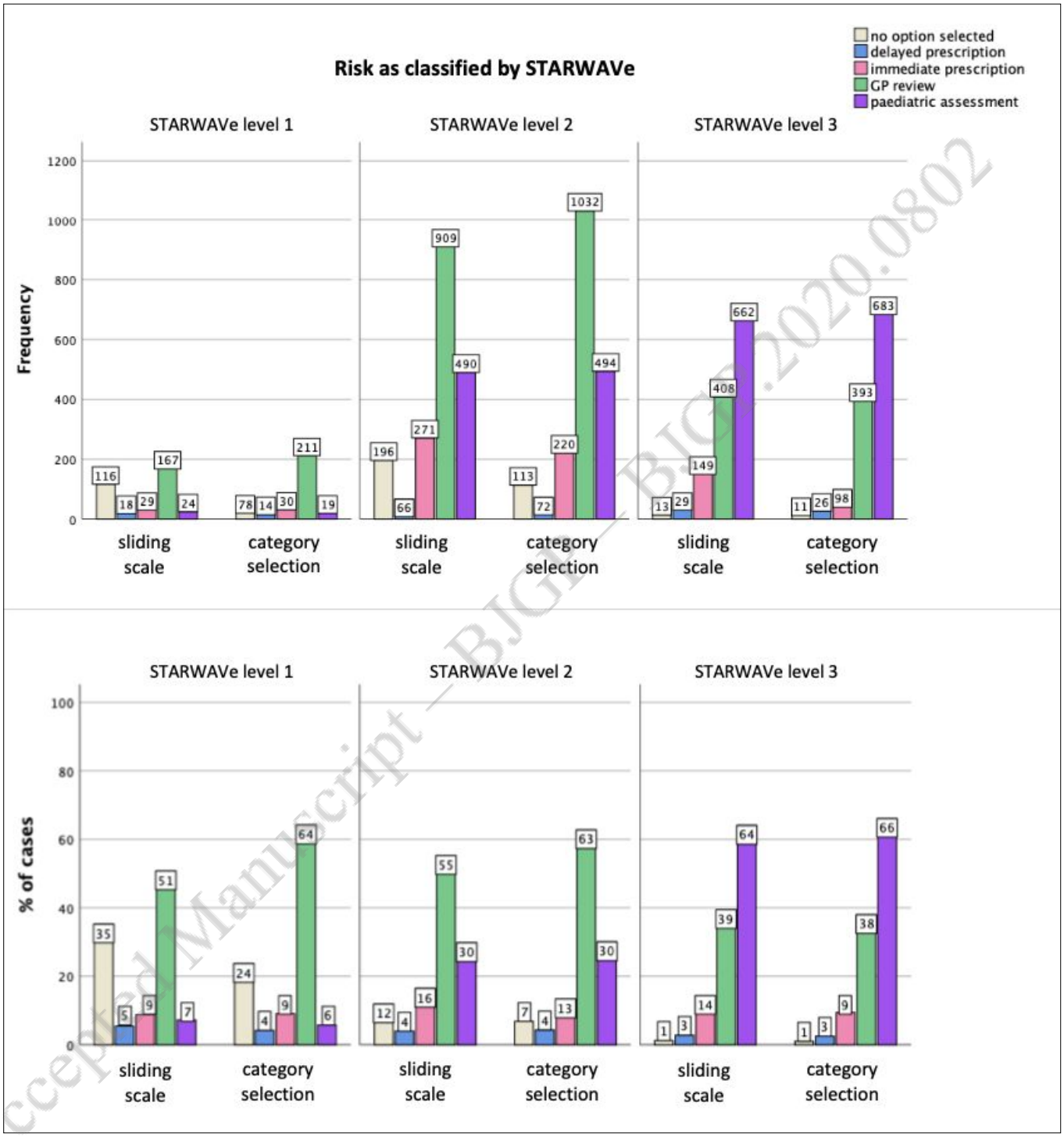


Figure 3. GPs' selections for patient management, by STARWAVe risk classification and risk response mode. The Figure displays the number (top) and proportion (bottom) of times that each option for patient management was chosen. Participants could select multiple options (or none) therefore percentages do not sum to 100. The total number of cases classified by STARWAVe as *level 1, level 2,* and *level 3* was 329, 1645, and 1034, respectively. The appropriate management strategies (according to STARWAVe) are: no action in *level 1* cases; no action or delayed prescription in *level*

2 cases; 24hr GP review and/or immediate prescription in *level 3* cases. Admission for paediatric assessment is not recommended by STARWAVe but is a reasonable strategy in more severe cases.

TABLE 1

Factor	Factor levels *	Factor range
Age	1: <2 years 0: ≥ 2 years	4 months to 6 years †
Illness duration	1: ≤ 3 days 0: >3 days	1 to 21 days ‡
Temperature	1: parent reports severe fever in the last 24 hours 0: none	N/A
Vomiting	1: parent reports moderate/severe vomiting in the last 24 hours 0: none	N/A
Current asthma	1: present 0: none	N/A
Inter/subcostal recession	1: present on examination 0: none	N/A
Wheeze	1: present on examination 0: none	N/A

Table 1. Manipulation of STARWAVe factors in the present study.
* Factor levels were based on the STARWAVe CPR, which assigns 1 point if the patient is aged <2 years; if illness duration is ≤ 3 days; if the parent reports severe fever in the last 24 hours; if the parent reports moderate or severe vomiting in the last 24 hours; if the patient has current asthma; if inter/subcostal recession is present on examination; and if wheeze is present on examination.
† Most vignettes (83%) were aged 1 to 6 years, which was the interquartile range in the prognostic cohort study that gave rise to STARWAVe. The full range in the STARWAVe cohort study was 3 months to 16 years.
‡ Most vignettes (73%) had an illness duration of 3 to 10 days, which was the interquartile range in the STARWAVe cohort study. The full range in the STARWAVe cohort study was 0 to 28 days.

TABLE 2

	Risk assessed on a sliding scale (*n*=94)	**Risk assessed via category selection (*n*=94)**	**Total (*N*=188)**
Grade Qualified GP GP trainee	 86 (92%) 8 (8%)	 81 (86%) 13 (14%)	 167 (89%) 21 (11%)
Number of years since GP qualification	*M*=10.2 (*SD*=8.9), range 0-35	*M*=11.3 (*SD*=9.4), range 0-38	*M*=10.7 (*SD*=9.1), range 0-38
Level of training (trainees) ST3 ST4	 6 (6%) 2 (2%)	 11 (12%) 2 (2%)	 17 (9%) 4 (2%)
Diploma in Child Health Yes No	 15 (16%) 79 (84%)	 18 (19%) 76 (81%)	 33 (18%) 155 (82%)
Member/Fellow of Royal College of Paediatrics & Child Health Yes No	 1 (1%) 93 (99%)	 2 (2%) 92 (98%)	 3 (2%) 185 (98%)
Self-reported confidence when assessing sick children	*M*=1.7 (*SD*=.6), range 0-3	*M*=1.9 (*SD*=.5), range 0-3	*M*=1.8 (*SD*=.6), range 0-3
Vignettes seen Set A, survey 1 first Set A, survey 2 first Set B, survey 1 first Set B, survey 2 first	 24 (26%) 23 (25%) 23 (25%) 24 (26%)	 24 (26%) 23 (25%) 24 (26%) 23 (25%)	 48 (26%) 46 (25%) 47 (25%) 47 (25%)

Table 2. Sample characteristics, per risk response mode and overall. Self-reported confidence when assessing sick children was measured on a 4-point scale (0="I seldom feel confident", 1="I feel confident sometimes", 2="I feel confident most of the time", 3="I always feel confident").

TABLE 3

STARWAVe factor	Risk assessments		Antibiotic prescriptions
	sliding scale *B* [95% CI]	category selection *OR* [95% CI]	*OR* [95% CI]
Duration (ascending)	0.09** [0.04, 0.13]	1.04* [1.01, 1.06]	1.19** [1.15, 1.23]
Temperature	2.53** [2.16, 2.90]	5.28** [4.09, 6.81]	7.18** [5.33, 9.68]
Age (ascending)	-0.26** [-0.35, -0.17]	0.88** [0.83, 0.93]	1.17** [1.10, 1.24]
Recession	5.29** [4.77, 5.81]	55.55** [39.28, 78.54]	0.45** [0.33, 0.62]
Wheeze	2.45** [2.15, 2.75]	6.53** [5.11, 8.35]	0.85 [0.68, 1.07]
Asthma	0.64** [0.40, 0.88]	2.07** [1.73, 2.47]	1.06 [0.85, 1.33]
Vomiting	1.66** [1.37, 1.95]	3.29** [2.65, 4.08]	0.82 [0.66, 1.03]

Table 3. Effect of STARWAVe risk factors on GPs' risk assessments and prescribing decisions.
**$p \leq 0.001$; *$p \leq 0.05$. Age and duration were treated as continuous in these models; when treated as binary, findings did not change (see Appendix 10).
Column 1 (risk assessments cast on a sliding scale): the model included random slopes for all seven STARWAVe factors.
Column 2 (risk assessments cast via category selection): a model that included random slopes for all seven STARWAVe factors would not converge, therefore we identified the random slopes that best improved model fit and added them to the model progressively until non-convergence occurred (see Appendix 11). The final model contained random slopes for recession, temperature, wheeze, and vomiting. Three factors violated the proportional odds assumption (duration, wheeze, vomiting): their effect reduced as risk assessments increased (see Appendix 12).
Column 3 (prescriptions): the model would not converge with random slopes for all STARWAVe factors, therefore we identified and included the random slopes that best improved model fit (see Appendix 13). The final model contained random slopes for all STARWAVe factors except age. When prescriptions were treated as a 3-category ordinal variable (0=no prescription, 1=delayed prescription, 2=immediate prescription), findings did not change (Appendix 14).

ADDITIONAL INFORMATION

Data availability: The data have been included in the submission and will be made publicly available on the Open Science Framework upon publication.

Funding: The study is funded by the National Institute for Health Research (NIHR) Imperial Patient Safety Translational Research Centre (http://www.imperial.ac.uk/patient-safety-translational-research-centre). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Ethical approval: The study has been approved by the Imperial Joint Research Compliance Office (ref 19IC5589) and the NHS Health Research Authority (ref 21/HRA/0958).

Patient and public involvement (PPI): The Imperial College London Research Partners Group (comprising patients, carers, and members of the public) gave feedback on the study design.

Competing interests: There are no competing interests.

Acknowledgements: The authors gratefully acknowledge infrastructure support from the NIHR Imperial Patient Safety Translational Research Centre, the NIHR Imperial Biomedical Research Centre and the NIHR-CRN.

REFERENCES

1. Department of Health. *Tackling Antimicrobial Resistance 2019-2024: The UK's Five-Year National Action Plan*; 2019. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/784894/UK_AMR_5_year_national_action_plan.pdf. Accessed July 25, 2020.
2. NHS England. *The NHS Atlas of Variation in Healthcare September 2015: Reducing Unwarranted Variation to Increase Value and Improve Quality*; 2015. http://www.rightcare.nhs.uk/atlas/RC_nhsAtlas3_HIGH_150915.pdf. Accessed August 16, 2016.
3. McCormick A, Fleming D, Charlton J, Office of Population Censuses and Surveys, Department of Health, Royal College of General Practitioners. *Morbidity Statistics from General Practice. Fourth National Study 1991-1992*. London: H.M.S.O.
4. Hay AD, Redmond NM, Turnbull S, et al. Development and internal validation of a clinical rule to improve antibiotic use in children presenting to primary care with acute respiratory tract infection and cough: a prognostic cohort study. *Lancet Respir Med* 2016;4(11):902-910.
5. Spinks A, Glasziou PP, Del Mar CB. Antibiotics for sore throat. *Cochrane Database Syst Rev* 2013(11).
6. Smith SM, Fahey T, Smucny J, Becker LA. Antibiotics for acute bronchitis. *Cochrane Database Syst Rev*, 2014(3).
7. Venekamp RP, Sanders SL, Glasziou PP, et al. Antibiotics for acute otitis media in children. *Cochrane Database Syst Rev* 2015(6).
8. Horwood J, Cabral C, Hay AD, et al. Primary care clinician antibiotic prescribing decisions in consultations for children with RTIs: a qualitative interview study. *Br J Gen Pract* 2016;66(644):e207-e213.
9. Lucas P, Cabral C, Hay AD, et al. A systematic review of parent and clinician views and perceptions that influence prescribing decisions in relation to acute childhood infections in primary care. *Scand J Prim Health Care* 2015;33(1):11-20.
10. Cabral C, Lucas PJ, Ingram J, et al. "It's safer to…" parent consulting and clinician antibiotic prescribing decisions for children with respiratory tract

infections: an analysis across four qualitative studies. *Soc Sci Med* 2015;136:156-164.

11. Guidance from the National Institute for Health and Care Excellence. Respiratory tract infections – antibiotic prescribing: prescribing of antibiotics for self-limiting respiratory tract infections in adults and children in primary care. http://www.nice.org.uk/guidance/cg69. Accessed May 23, 2019.
12. Turnbull SL, Redmond NM, Lucas P, et al. The CHICO (Children's Cough) Trial protocol: a feasibility randomised controlled trial investigating the clinical and cost-effectiveness of a complex intervention to improve the management of children presenting to primary care with acute respiratory tract infection. *BMJ Open* 2015;5(9):e008615.
13. Chiang PP-C, Glance D, Walker J, et al. Implementing a QCancer risk tool into general practice consultations: an exploratory study using simulated consultations with Australian general practitioners. *Br J Cancer* 2015;112(s1):S77.
14. Nurek M, Delaney BC, Kostopoulou O. Risk assessment and antibiotic prescribing decisions in children presenting to UK primary care with cough: a vignette study. *BMJ Open* 2020;10(7):e035761.
15. Sirota M, Round T, Samaranayaka S, et al. Expectations for antibiotics increase their prescribing: causal evidence about localized impact. *Health Psychology* 2017;36(4):402.
16. StataCorp L. Stata multilevel mixed-effects reference manual: release 13. College Station, TX;2013.
17. Williams R. Understanding and interpreting generalized ordered logit models. *J Math Sociol* 2016;40(1):7-20.
18. Williams R. Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *Stata J* 2006;6(1):58-82.
19. Schulte-Mecklenbeck M, Kühberger A, Ranyard R. *A handbook of process tracing methods for decision research: A critical review and user's guide.* New York and Hove: Psychology Press; 2010.
20. Nurek M, Kostopoulou O. What you find depends on how you measure it: Reactivity of response scales measuring predecisional information distortion in medical diagnosis. *PLOS One* 2016;11(9):e0162562.

21. Mangione-Smith R, McGlynn EA, Elliott MN, et al. The relationship between perceived parental expectations and pediatrician antimicrobial prescribing behavior. *Pediatrics* 1999;103(4):711-718.
22. Dempsey PP, Businger AC, Whaley LE, et al. Primary care clinicians' perceptions about antibiotic prescribing for acute bronchitis: a qualitative study. *BMC Fam Pract* 2014;15(1).
23. Jakobsen KA, Melbye H, Kelly MJ, et al. Influence of CRP testing and clinical findings on antibiotic prescribing in adults presenting with acute cough in primary care. *Scand J Prim Health Care* 2010;28(4):229-236.
24. Petursson P. GPs' reasons for "non-pharmacological" prescribing of antibiotics: a phenomenological study. *Scand J Prim Health Care* 2005;23(2):120-125.
25. Butler CC, Rollnick S, Pill R, et al. Understanding the culture of prescribing: qualitative study of general practitioners' and patients' perceptions of antibiotics for sore throats. *BMJ* 1998;317(7159):637-642.
26. Borek AJ, Campbell A, Dent E, et al. Implementing interventions to reduce antibiotic use: a qualitative study in high-prescribing practices. *BMC Fam Pract* 2021;22(25):1-11.