APPENDIX A. METHODOLOGYCAL AND MATHEMATICAL FOUNDATIONS

A1. Anchor scenarios

Specifically, the set of red scenarios, Ω_R, are defined as follows:

$$\Omega_R = \{S_j \in \Omega_U \ / \nexists \ S_r \in \Omega_U \ satisfying \ S_r < S_j\}$$

where

$$\Omega_U = \{S_j \in \Omega \ / \ f(S_j) \geq U_R\},$$

and U_R is a high level of stress, for example, $U_R \approx 90$.

Similarly, green scenarios are obtained by combining different factor levels provided by ED workers that do not contribute to high levels of stress (low priority patients, waiting time targets achieved, a small amount of patients, etc.). The set of green scenarios, Ω_G, are defined as follows:

$$\Omega_G = \{S_j \in \Omega_{\mathrm{L}} \ / \nexists \ S_g \ \in \Omega_L \ satisfying \ S_g > S_j\}$$

where

$$\Omega_L = \{S_j \in \Omega \ / \ f(S_j) \leq U_G\},$$

and $\mathrm{U_G}$ is a low level of stress, for example, $\mathrm{U_G} \approx 10$.

A2. Homogenization of experts' answers in a common scale

Let $y_i^*(S_j) = y_{ij}^*$ be the stress score for scenario S_j provided by physician i. In case the range of values $\{y_{ij}^* \ / \ j = 1, \dots, 16\}$ greatly differs from the total range [0, 100], a transformation $g(y_{ij}^*) = y_{ij}$ is needed such that the range of values $\{y_{ij} \ / \ j = 1, \dots, 16\}$ is similar to [0, 100]. Observe that each physician has to assess green scenarios (with minimum stress associated) and red scenarios (with maximum stress associated).

This transformation should preserve the ordering of scenarios and the ratio of differences in stress among them. Any non-decreasing transformation preserves the ordering of the scenarios. In addition, the second condition leads us to a linear transformation.

That is, for any given two scenarios S_u, S_v and any scenario S_j, it is imposed that

$$\frac{y_{ij}^* - y_{iv}^*}{y_{iu}^* - y_{iv}^*} = \frac{y_{ij} - y_{iv}}{y_{iu} - y_{iv}}$$

From which a linear relationship between y_{ij} and y_{ij}^* is readily obtained:

$$\boldsymbol{y_{ij}} = \frac{y_{iu} - y_{iv}}{y_{iu}^* - y_{iv}^*} \times \boldsymbol{y_{ij}^*} + \left(y_{iv} - \frac{y_{iu} - y_{iv}}{y_{iu}^* - y_{iv}^*} \times y_{iv}^* \right)$$

$$y_{ij} = g(y_{ij}^*) = a_i + b_i \times y_{ij}^*$$

where

$$a_i = \left(y_{iv} - \frac{y_{iu} - y_{iv}}{y_{iu}^* - y_{iv}^*} \times y_{iv}^* \right)$$

$$b_i = \left(\frac{y_{iu} - y_{iv}}{y_{iu}^* - y_{iv}^*} \right)$$

The scenarios providing the pairs $(y_{iu}^*, y_{iu}), (y_{iv}^*, y_{iv})$, which determine the parameters of the linear transformation, are those introduced as anchors in the questionnaire. U_R is the expected stress induced by a red scenario. This value is estimated by the trimmed mean of the scores provided by physicians for those red scenarios:

$$y_i^R = \max\{y_{ir}/S_r \in \Omega_R\}$$

$$\{y_{[j]}^R\}_{j=1}^N \textit{the ascending ordered set } \{y_i^R\}_{i=1}^N$$

$$\frac{1}{n-2p} \sum_{j=p+1}^{n-p} y_{[j]}^R = \widehat{U}_R$$

Furthermore, the confidence interval (CI) for U_R is calculated, and those physicians, whose scores for red scenarios are below the left limit and gave scores on the lower side of the scale – e.g. because they have a higher stress threshold than their colleges – need to be rescaled according to the linear transformation.

U_G is the expected stress induced by a green scenario, and its value is estimated by the trimmed mean of the scores provided by physicians for those green scenarios $\widehat{U}_G$, similar to $\widehat{U}_R$. The CI for U_G is also calculated and those physicians, whose scores for green scenarios are above the right limit, need to be rescaled.

When a physician i with a high stress threshold uses only the low side of the stress scale and their scores for red scenarios, y_i^R, are below the CI calculated (case 1), then the pairs for the transformation are $\left(y_i^R, \widehat{U}_R\right)$ and (y_i^G, y_i^G). In the opposite case, when a physician i has all their scenario scores in the upper side of the stress scale (case 2) and y_i^G is above the CI calculated for U_G, the pairs for the transformation are (y_i^R, y_i^R) and $(y_i^G, \widehat{U}_G)$. Finally, if a physician i has all their values concentrated on the middle of the scale (case 3), then the pairs for the linear transformation are $\left(y_i^R, \widehat{U}_R\right)$ and $(y_i^G, \widehat{U}_R)$.

If a physician i has spread all their values over the stress scale (case 4), these do not need to be rescaled. These four homogenization cases are represented in Figure A. 1, where PS scores are the scores provided by physicians on their personal scale, and CS scores are the physicians' scores on the common scale.

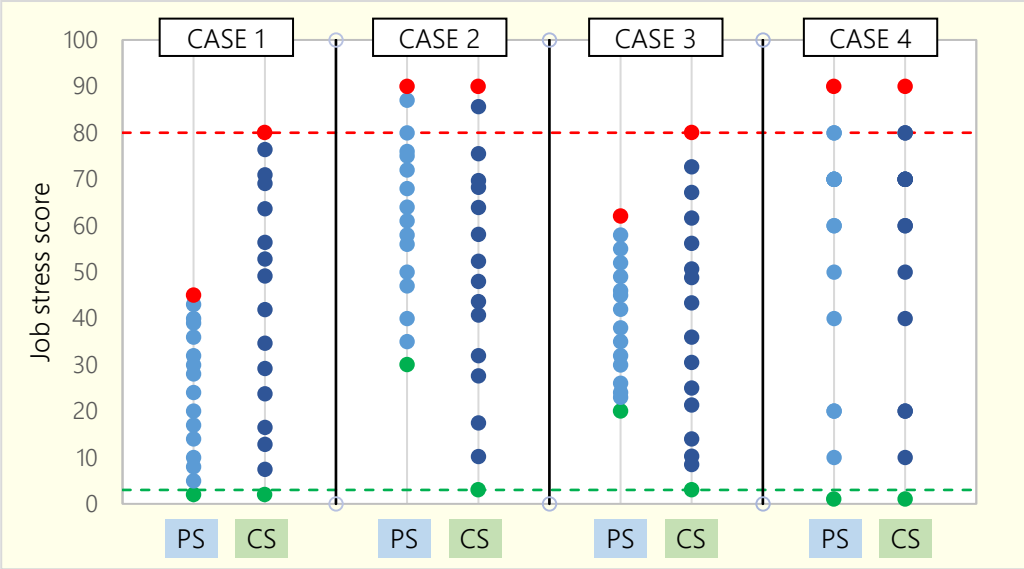


Figure A. 1. Cases of homogenization for physicians' scores.

A3. Coherence and consistency analysis for each expert's answers

Coherence. To analyse the coherence of a physician, it is necessary to introduce the concept of dominance between scenarios:

A scenario S_1 defined by the vector of stress variables $\{X_{1,1}, \dots, X_{1,11}\}$ dominates over a scenario $S_2 \equiv \{X_{2,1}, \dots, X_{2,11}\}$, represented by $S_1 > S_2$, if and only if

$$X_{1k} \geq X_{2k} \; \forall \; k = 1, \dots, 11 \; and \; \exists \; j \; s.t. X_{1j} > X_{2j}$$

A physician is coherent assessing scenario S_1 and S_2 if $S_1 > S_2 \Leftrightarrow y_i(S_1) > y_j(S_2)$.

A coherence index, CoI, similar to the Kendall's tau-a is defined by taking into account the pairs of scenarios with a dominance relationship which is coherently and incoherently assessed, denoted by D_c and D_D, respectively:

$$CoI = \frac{D_c - D_D}{D_c + D_D}$$

Physicians whose CoI are below a certain threshold are excluded.

Consistency. Many researchers in medicine, biology, engineering, etc. need measures of agreement aimed to assess the reproducibility of judgments. The concept of inter-rater reliability expresses our need of quality for measurement, in terms of concordance of judgments - as this study looks for a consensus among physicians.

Most of them propose the Kappa Statistic, a statistic that indicates the degree of agreement from nominal or ordinal assessments. However, when there are ordinal ratings, Kendall's coefficients are more appropriate statistics to determine association as they take ordering into consideration.

We check the consistency of a physician by comparing his/her answers with those physicians that answered the same questionnaire (denoted as group of physicians). Thus, "Kendall's correlation coefficient tau-b" could be more appropriate to use as it measures association between two ordinal variables, each appraiser (one physician) with the known "standard" (the consensus from the rest of the group).

Now, the question of how to define the "standard" arises. One possibility is to create the standard by averaging the scores provided by other physicians or by selecting the median answer or other statistic. However, these values could not represent the majority's opinion of the group. Suppose three raters provide (10, 10, 15) to S_1 and (12, 12, 6) to S_2, the majority agrees that S_2 is more stressful than S_1, but $\overline{y(S_1)} < \overline{y(S_2)}$. To avoid these undesirable situations, we define the standard directly from a voting system. One scenario is considered more stressful than the other when the majority of the group considers it so. If there is a tie, then we have a "indecisiveness" situation.

A group of n physicians is denoted as $A = \{D_1, \dots, D_i, \dots, D_n\}$, the set of m scenarios forming the questionnaire answered by the group A as $\Omega_A = \{S_1, \dots, S_j, \dots, S_m\}$ $(\Omega_A \subseteq \Omega)$ and the stress score of a physician i for scenario j $Y_i(S_j) = Y_{ij}$. We construct the matrix, M_i, which indicates above the main diagonal the stress comparisons between scenarios made by physician i.

$$M_i = [m_i(j,k)] \;\; such\ that \;\; \forall j \geq k \; m_i(j,k) = 0 \;\; and \;\; \forall j < k \;\; m_i(j,k) = \begin{cases} 1\ if\ Y_{ij} < Y_{ik} \\ 0\ if\ Y_{ij} = Y_{ik} \\ -1\ if\ Y_{ij} > Y_{ik} \end{cases}$$

Another matrix Q_A is defined to reflect the consensus of group A of physicians about their stress comparisons between scenarios.

$$Q_A = [q_A(j,k)] \;\; such\ that \;\; q_A(j,k) = \begin{cases} 1\ if\ (\sum_{i \in A} m_i(j,k)) > 0 \\ 0\ if\ (\sum_{i \in A} m_i(j,k)) = 0 \\ -1\ if\ (\sum_{i \in A} m_i(j,k)) < 0 \end{cases} \quad \forall\, i \neq h$$

The agreements of a physician h with the rest of physicians in his/her group A are stored in a matrix $G_{A(h)}$, defined from the matrices M_i and $Q_{A_{-h}}$, where A_{-h} denotes the set A minus the physician h ($A_{-h} = A - \{h\}$):

$$G_{A(h)} = [g_{A(h)}(j,k)] \;\; such\ that \;\; \forall j \geq k \;\; g_{A(h)}(j,k) = 0 \;\; and$$

$$\forall j < k \quad g_{A(h)}(j,k) = \begin{cases} 1\ if\ \ m_i(j,k) = q_{A_{-h}}(j,k), & concordance \\ 0\ if\ \ q_{A_{-h}}(j,k) = 0\ and\ \ m_i(j,k) \neq 0, & indecisiveness \\ -1\ if\ \ q_{A_{-h}}(j,k) = -\,m_i(j,k), & discordance \end{cases}$$

The values $g_{A(h)}(j,k)$ reflect three situations between a physician h and the rest of the group:

Concordance: physician and the rest of the group assigned the same order to a pair of scenarios S_j, S_k from most stressful to less stressful.

Indecisiveness: one half and the other half of the group members' assigned the opposite order to a pair of scenarios S_j, S_k from most stressful to less stressful. Physician h would break the tie among the rest of the group physicians.

Discordance: physician and the rest of the group assigned the opposite order to a pair of scenarios S_j, S_k from most stressful to less stressful.

Finally, the Kendall's tau-b is adapted to consider the three possible cases. The consistency with the group index, CGI, taking into consideration the number of concordances ($C = \sum_{j,k} 1_{\{g_{A(h)}(j,k)=1\}}$), discordances ($D = \sum_{j,k} 1_{\{g_{A(h)}(j,k)=-1\}}$), and indecisiveness ($I = \sum_{j,k} 1_{\{g_{A(h)}(j,k)=0\}}$) in the matrix $G_{A(h)}$ is defined as:

$$CGI_{A(h)} = \frac{(C - D)}{(C + D + I)}$$

An "inconsistent rater" – who should be excluded for the study – is a rater whose CGI is below a certain limit $L_l \in \mathbb{R}$. In this study, we consider $L_l = 0.25$.

APENDIX B. Questionnaires for eliciting expert opinion

B1. Instructions sheet for completion of the questionnaire

upna Gobierno de Navarra

Research Group DECYL

ASSESSMENT OF STRESS DUE TO INSTANTANEOUS WORKLOAD

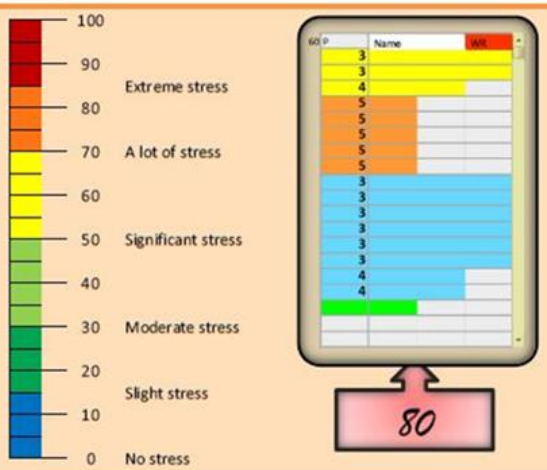
The following survey evaluates the **stress** perceived by physicians due to their assigned **workload at a given time**, which is called the **instantaneous workload**. An instantaneous workload is considered the **set of pending patients a doctor has been assigned simultaneously** – distinguishing their priority, as well as the **medical attention phase** of the patients– and his/her training responsibility.

To carry out this assessment, physicians will be presented several scenarios of different pending workloads associated to a physician in the same way as the patients' portfolio of the emergency department computer. This board lets physicians be aware of all the pending patients they have been assigned to them as patients are triaged and then immediately assigned to a specific physician as they arrive to the Emergency Department. The following colour code is used:

Priority	Name WoutR / WR
3, 4, 5	Patients waiting to be seen for the first time.
3, 4, 5	Patients waiting to be seen for the first time, they have exceeded the waiting time limit.
3, 4, 5	Patients who have already been seen by the doctor, who have requested medical tests. These patients are still inside the ED system waiting for results and being discharge.
3, 4, 5	Patients who have receive the medical discharge, but remain in the emergency department waiting to be transferred to the hospital, home, etc. In case of these patients to get worse, the assigned physician is still responsible for your care

Patients who have been discharged and have left the system are not considered as pending workload.

What follows is a set of scenarios which show several patients a doctor has been assigned at a specific moment of his/her work-shift, and whether ***the shift*** he/she is working does or does not ***involve the supervision of a first-year resident*** (indicated in the red colour at the top right corner of the computer screen represented above. The experts should assess the perceived stress associated to each of the scenarios. This will involve that experts give a score between 0 and 100, helped by the qualitative scale described below.

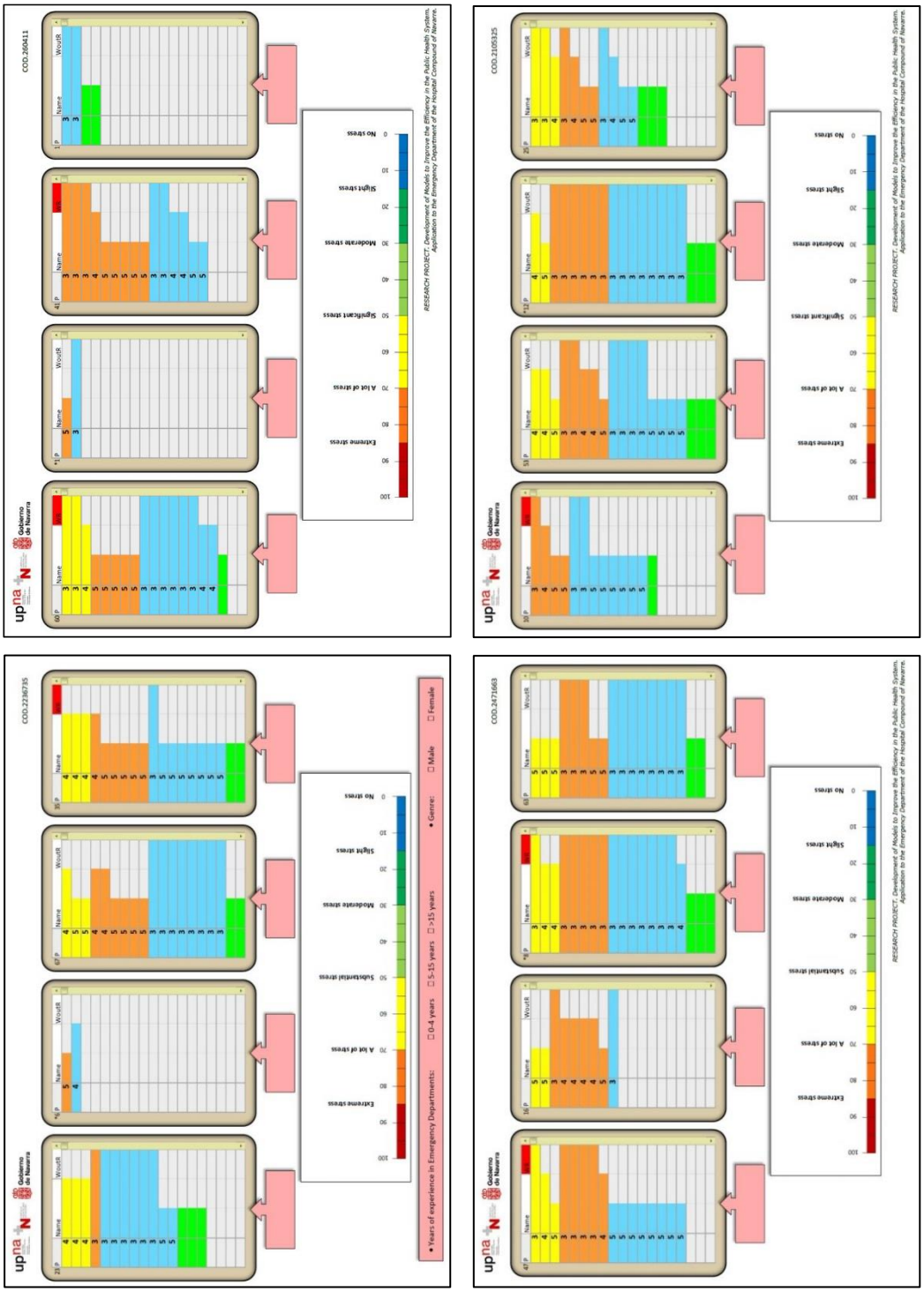


Scenario:
The physician has been assigned three patients whose waiting time limit has been exceeded (two of priority 3 and one of priority 4). In addition, there are also five patients with minimum severity index (5) to be seen, and eight patients (six with priority 3) to receive a second consultation and be discharged. Moreover, the physician has a patient who has got his/her ED process finished but who is waiting to be transferred to another hospital or home. Finally, he has to supervise a resident.

Physician Stress Assessment:
If the physician considers that the stress produced by the workload assigned in his/her patients' portfolio is very high, he/she could assess the stress situation with a score of 80.

***** Note: There are no correct or incorrect ratings. You are asked to provide a subjective assessment of the stress due to workload assigned to the physician at a given point in the work shift.***

RESEARCH PROJECT. Development of Models to Improve the Efficiency in the Public Health System. Application to the Emergency Department of the Hospital Compound of Navarre.

B2. Questionnaire Example.

In this appendix, we show one of the six questionnaires designed. They only differ in the set of scenarios provided in each of the four cards to be assessed in terms of stress.