# ONLINE APPENDIX

This document includes all supplementary material to the paper "Effects of incentive framing on performance and effort: Evidence from a medically framed experiment", by Mylene Lagarde and Duane Blaauw.

## Content

## Appendix A: Additional Tables and Figures

**Table A1: Characteristics of respondents**

| | Control | | Loss | | Gain | | Kruskal-Wallis test p-value |
|---|---|---|---|---|---|---|---|
| Age | 23.38 | (1.56) | 23.32 | (1.56) | 23.57 | (1.87) | 0.880 |
| Female | 0.63 | (0.49) | 0.60 | (0.49) | 0.68 | (0.47) | 0.634 |
| White | 0.18 | (0.39) | 0.18 | (0.39) | 0.18 | (0.39) | 1.000 |
| Black | 0.67 | (0.48) | 0.65 | (0.48) | 0.62 | (0.49) | 0.844 |
| In bottom 20% of class | 0.30 | (0.46) | 0.18 | (0.39) | 0.28 | (0.45) | 0.286 |
| Diagnosis knowledge test score | 6.72 | (2.41) | 7.12 | (2.63) | 7.07 | (2.28) | 0.698 |
| Extravert | 0.09 | (0.96) | -0.07 | (1.11) | 0.03 | (1.07) | 0.748 |
| Intellectual | 0.03 | (1.02) | 0.08 | (0.98) | -0.02 | (1.13) | 0.238 |
| Neurotic | 0.07 | (0.94) | 0.07 | (1.02) | -0.12 | (1.07) | 0.709 |
| Conscientious | 0.16 | (1.00) | -0.11 | (0.99) | -0.06 | (1.05) | 0.787 |
| Agreeable | 0.12 | (1.12) | -0.05 | (0.95) | -0.03 | (1.02) | 0.515 |
| *N* | 60 | | 60 | | 60 | | |

Note: The table reports the mean for each variable, with standard deviations reported in parentheses. To measure participants' Big 5 personality traits, we used the International English Big-Five Mini-Markers, a validated tool comprising 40 single adjective personality descriptors, designed to be used in an international context where respondents may not all be native English-speakers. Each of the five traits is measured separately as the sum of all relevant questions (inversely coded for negative adjectives), and then normalized into z-scores. The last column reports the p-value of a Kruskal-Wallis test of equality across the three treatments.

**Table A2: Descriptive statistics of performance measures, by treatment**

| | Control | | | Gain | | | Loss | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | Median | SD | Mean | Median | SD | Mean | Median | SD |
| **Data entry** | | | | | | | | | |
| # of data sheets done | 5.95 | 6.00 | 2.00 | 6.48 | 6.00 | 1.85 | 6.00 | 6.00 | 1.57 |
| Total # of correct entries (performance) | 96.93 | 94.50 | 41.24 | 116.62 | 112.00 | 44.28 | 117.83 | 116.50 | 34.84 |
| Total # of attempted entries | 119.27 | 110.00 | 45.77 | 128.20 | 132.00 | 41.76 | 120.58 | 118.50 | 35.53 |
| # of correct/attempted entries (accuracy) | 0.89 | 0.99 | 0.28 | 0.93 | 0.99 | 0.19 | 0.98 | 0.99 | 0.06 |
| Time spent on data entry (effort persistence) | 390.01 | 388.52 | 69.80 | 419.82 | 429.60 | 54.62 | 394.47 | 398.13 | 53.28 |
| Time per attempted data entry | 3.73 | 3.63 | 1.41 | 3.72 | 3.28 | 1.71 | 3.57 | 3.45 | 1.27 |
| Time per correct data entry | 6.97 | 4.12 | 13.52 | 4.36 | 3.48 | 2.62 | 3.67 | 3.45 | 1.42 |
| **Diagnosis** | | | | | | | | | |
| Total # of correct diagnoses (performance) | 3.18 | 3.00 | 1.53 | 3.48 | 3.50 | 1.76 | 3.35 | 3.00 | 1.52 |
| Total # of attempted diagnoses | 5.30 | 5.00 | 2.12 | 5.67 | 6.00 | 2.10 | 5.28 | 5.00 | 1.67 |
| # of correct/attempted diagnoses (accuracy) | 0.67 | 0.73 | 0.29 | 0.64 | 0.67 | 0.26 | 0.66 | 0.67 | 0.27 |
| Time spent on diagnosis (effort persistence) | 144.65 | 154.11 | 62.13 | 122.04 | 122.86 | 51.28 | 149.84 | 162.60 | 52.24 |
| Time per attempted diagnosis | 33.58 | 30.76 | 24.07 | 25.47 | 19.98 | 19.28 | 31.02 | 30.62 | 14.06 |
| Time per correct diagnosis | 52.36 | 47.95 | 28.85 | 43.46 | 34.91 | 35.35 | 52.64 | 46.71 | 32.69 |
| *N* | 60 | | | 60 | | | 60 | | |

Note: The table presents descriptive data of performance measures. Performance in data entry (diagnosis identification) is equal to the number of correct entries (diagnoses) made by the participant. Effort intensity is measured by the subject accuracy in data entry (diagnosis identification), calculated as the ratio of the number of correct entries (diagnoses) over of all attempted entries (diagnoses). Effort persistence in an activity is defined as the time spent on the activity over the entire period of work (in seconds). Note that the mean time per attempted data entry [diagnosis] is calculated as the average of individual ratios of total time per attempted data entries [diagnoses] over total number of attempts.

**Table A3: Performance and effort channels in data entry in period 2**

| | Performance (# of correct entries) | | Effort persistence (time spent) | | Effort intensity (accuracy) | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Gain | 33.000*** | 26.736** | 36.032** | 41.068** | 0.045 | 0.044 |
| | (11.521) | (10.883) | (15.373) | (16.458) | (0.058) | (0.063) |
| Loss | 38.133*** | 26.247** | 26.083* | 31.766* | 0.152** | 0.129** |
| | (11.521) | (10.687) | (15.373) | (16.161) | (0.058) | (0.062) |
| Mean in control group (SD) | 113.2 | | 413.3 | | 0.825 | |
| | (47.33) | | (65.72) | | (0.295) | |
| Observations | 90 | 90 | 90 | 90 | 90 | 90 |
| Individual controls | No | Yes | No | Yes | No | Yes |
| $R^2$ | 0.129 | 0.378 | 0.063 | 0.141 | 0.077 | 0.130 |

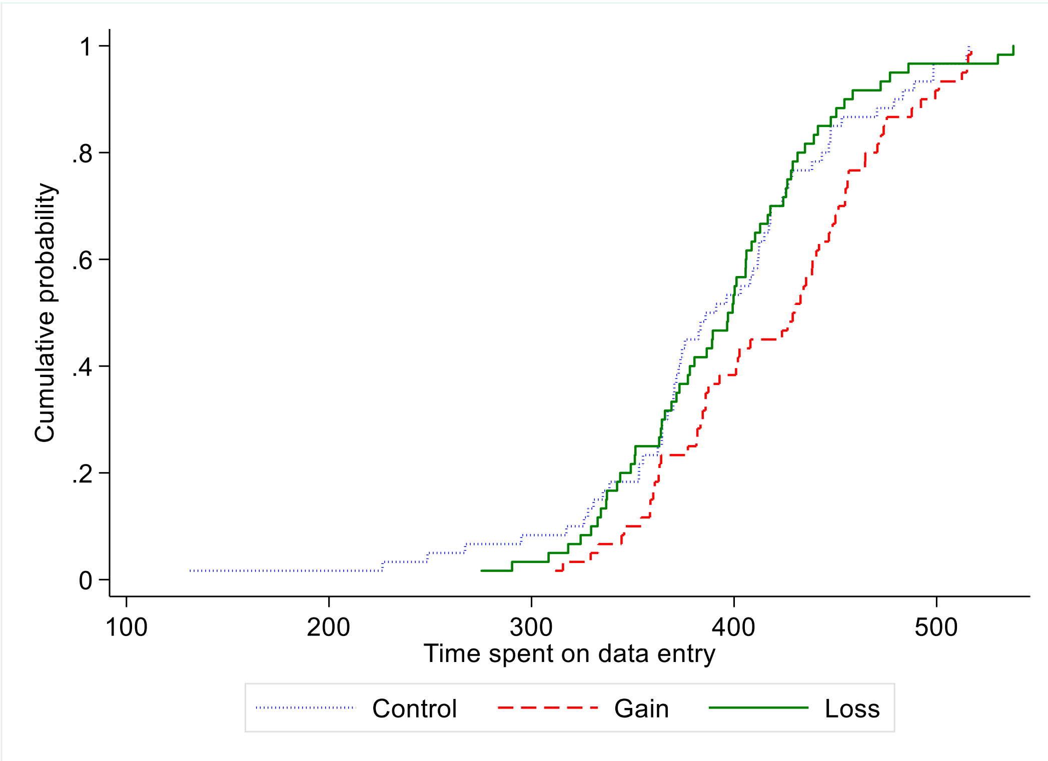
*Note:* the sample is restricted to participants who were randomised to version B of the experiment in period 2, meaning they work under exactly the same conditions as they do in period 1. The dependent variable in columns 1-2 is equal to the total number of correct entries made by the participant over the 10-mn period of work. In columns 3.4, the dependent variable is defined as the time spent on data entry over the entire period of work (in seconds). In columns 5-6, the dependent variable (effort intensity) is measured by the subject's accuracy in data entry, calculated as the ratio of total number of correct entries over all attempted entries during the period. Individual controls include age, gender, ethnicity, grade obtained the previous year, knowledge of blood test result interpretation and personality traits (Big 5 inventory). Standard errors in parentheses are clustered at the individual level. *** p < .001 ** p < .01, * p < .05

**Table A4: Performance and effort channels in diagnosis identification in period 2**

| | Performance (# of diagnoses correctly identified) | | Effort persistence (time spent) | | Effort intensity (accuracy) | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Gain | 0.633 | 0.146 | -36.775** | -39.642** | -0.029 | -0.072 |
| | (0.659) | (0.583) | (14.533) | (15.544) | (0.082) | (0.080) |
| Loss | -0.167 | -0.928 | -26.508* | -31.246** | -0.041 | -0.110 |
| | (0.659) | (0.572) | (14.533) | (15.263) | (0.082) | (0.079) |
| Mean in control group (SD) | 3.567 | | 129.4 | | 0.551 | |
| | (2.515) | | (62.52) | | (0.336) | |
| Observations | 90 | 90 | 90 | 90 | 90 | 90 |
| Individual controls | No | Yes | No | Yes | No | Yes |
| $R^2$ | 0.018 | 0.387 | 0.073 | 0.151 | 0.003 | 0.232 |

*Note*: the sample is restricted to participants who were randomised to version B of the experiment in period 2, meaning they work under exactly the same conditions as they do in period 1. The dependent variable in columns 1-2 is equal to the total number of correct diagnoses identified by the participant over the 10-mn period of work. In columns 3.4, the dependent variable is defined as the time spent on diagnosis identification over the entire period of work (in seconds). In columns 5-6, the dependent variable (effort intensity) is measured by the subject's accuracy in diagnosis identification, calculated as the ratio of total number of correct diagnoses over all attempted ones during the period. Individual controls include age, gender, ethnicity, grade obtained the previous year, knowledge of blood test result interpretation and personality traits (Big 5 inventory). Standard errors in parentheses are clustered at the individual level. *** p < .001 ** p < .01, * p < .05

**Figure A1: Distribution of time spent on data entry, by treatment**



## Appendix B: Full experimental design

In the experiment, subjects took part in one of four treatments: control, gain, loss or tournament. Allocation of subjects to the first three treatments was done at random over 11 experimental sessions (see Appendix D), while participants in the tournament treatment consisted of all individuals who showed up at three separate sessions. The present paper therefore focuses on the comparison of the two incentive framings (gain and loss) and the control group. Two reasons led us to exclude the Tournament arm from the present paper:

- *Internal validity concerns due to implementation issues*: the experimental procedures were slightly different for the tournament treatment, compared to the other three treatments. While the allocation of subjects to the three treatments presented here was done at random over 11 experimental sessions, the tournament treatment was organised in three separate sessions, where all individuals who showed up took part in a 'tournament' session. Hence, in the absence of random assignment to these sessions, we cannot rule out that there are unobserved differences between the subjects who chose these sessions compared to the other subjects.
- *Conceptual issues*: while the loss and gain treatments were clearly only differing in terms of their framing, the fourth treatment used a completely different design (all individuals were given a fixed remuneration of R100 and they were told that the top 20% of performers would earn an additional R25). This choice was driven by intellectual curiosity rather than a clear set of hypotheses informed by theory. Hence it would have been difficult to make sense of the results.

In the experiment, participants take part in two consecutive periods of work of 10 minutes each. At the beginning of each period, instructions given on the screen provide detailed information about payment modalities (see Appendix E).

- Period 1 acts as a baseline: we observe subjects effort under the three main incentives, as described in the body of the text ; there is no difference between versions A and B of the treatments.
- In period 2, for participants randomised to version "A" could earn a private reward worth R7.50 per correct diagnosis made. Participants randomised to version B work under identical conditions as they did in period 1.

Table B1 below provides a summary of our experimental design.

**Table B1: Experimental design**

| Treatment | Period 1 | | Version | Period 2 | |
|---|---|---|---|---|---|
| | **Data entry** | **Diagnosis** | | **Data entry** | **Diagnosis** |
| **CONTROL** | No incentive | No incentive | A | No incentive | Reward |
| | | | B | | No reward |
| **GAIN** | Reward | No incentive | A | Reward | Reward |
| | | | B | | No reward |
| **LOSS** | Fine | No incentive | A | Fine | Reward |
| | | | B | | No reward |
| **TOURNAMENT** | Reward for 20% top performers | No incentive | A | Reward for 20% top performers | Reward |
| | | | B | | No reward |

*Note*: We abstract away from all social incentives (donations to charities): 'reward' or f'ine' here refers to monetary incentive for the subjects themselves. No incentive means no incentive for the activity for the subjects themselves.

## Appendix C: Details of the real effort task

### Rationale

The experiment involved a real effort task that was framed in a medical context and constructed to reproduce some salient characteristics of the provision of medical care, especially the fact that individuals are expected to perform several tasks, some of which that are repetitive and mundane, and others that are more cognitively engaging.

### Entering data from laboratory reports

In each period participants were handed a pile of 10 hardcopy laboratory test reports, as shown in Figure C1 below.

**Figure C1: Example of a laboratory report**

LABORATORY REPORT

**REFERENCE NUMBER** 421

Patient age: 29

**HAEMATOLOGY AND BIOCHEMISTRY RESULTS**

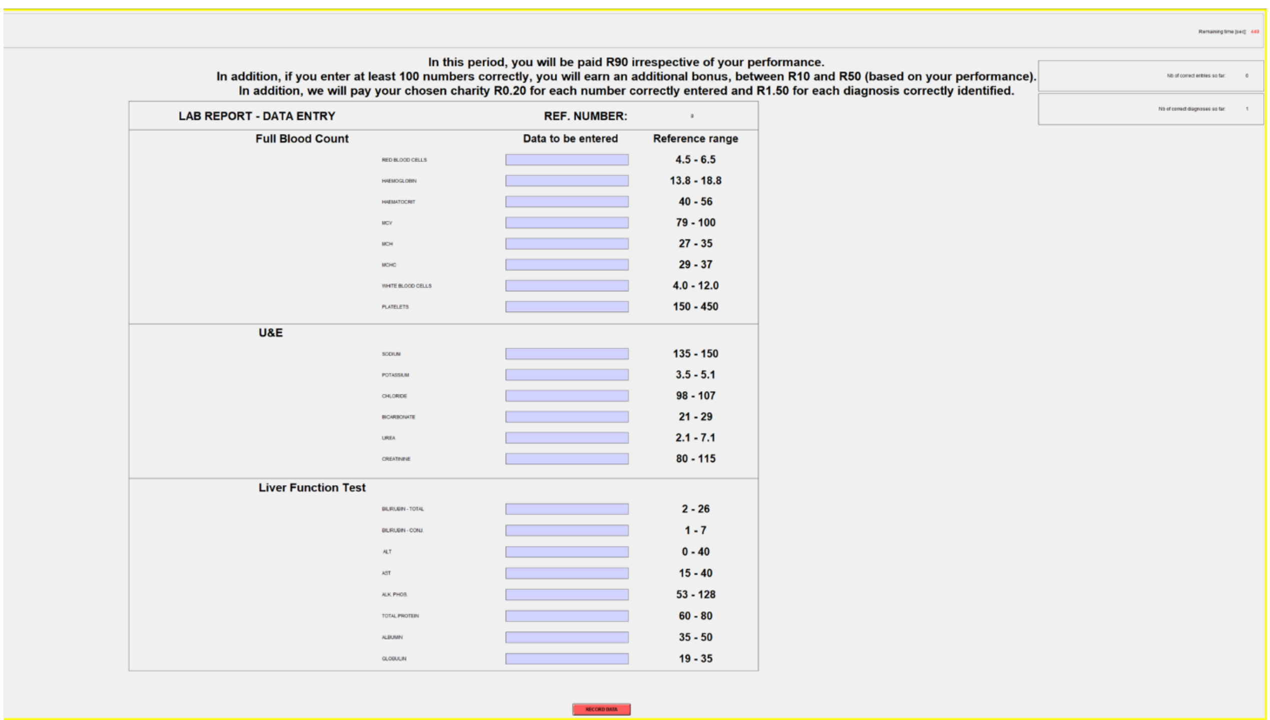
| Test | Result | Units | Reference Range |
|---|---|---|---|
| Full Blood Count | | | |
| RED BLOOD CELLS | **5.8** | x $10^{12}$/L | 4.5 - 6.5 |
| HAEMOGLOBIN | **15.2** | g/dL | 13.8 - 18.8 |
| HAEMATOCRIT | **47.2** | % | 40 - 56 |
| MCV | **89.8** | fL | 79 - 100 |
| MCH | **27.5** | pg | 27 - 35 |
| MCHC | **35.1** | g/dL | 29 - 37 |
| WHITE BLOOD CELLS | **7.2** | x $10^{9}$/L | 4.0 - 12.0 |
| PLATELETS | **317** | x $10^{9}$/L | 150 - 450 |
| U&E | | | |
| SODIUM | **142.5** | mmol/L | 135 - 150 |
| POTASSIUM | **3.7** | mmol/L | 3.5 - 5.1 |
| CHLORIDE | **103.2** | mmol/L | 98 - 107 |
| BICARBONATE | **23.5** | mmol/L | 21 - 29 |
| UREA | **6.5** | mmol/L | 2.1 - 7.1 |
| CREATININE | **95.8** | µmol/L | 80 - 115 |
| Liver Function Test | | | |
| BILIRUBIN - TOTAL | **6.2** | µmol/L | 2 - 26 |
| BILIRUBIN - CONJUGATED | **6.0** | µmol/L | 1 - 7 |
| ALT | **10.9** | IU/L | 0 - 40 |
| AST | **16.5** | IU/L | 15 - 40 |
| ALKALINE PHOSPHATASE | **127.6** | IU/L | 53 - 128 |
| TOTAL PROTEIN | **66.1** | g/L | 60 - 80 |
| ALBUMIN | **35.2** | g/L | 35 - 50 |
| GLOBULIN | **21.7** | g/L | 19 - 35 |

Each report presented 22 different individual test results from a Full Blood Count, a Urea & Electrolytes analysis, and a Liver Function Test:

- The Full Blood Count included 8 individual results: red blood cell count (RBC), haemoglobin (HGB), haematocrit (HCT), mean corpuscular volume (MCV), mean corpuscular haemoglobin (MCH), mean corpuscular haemoglobin concentration (MCHC), white blood cells count (WBC), platelet count (PLT).
- The Urea & Electrolytes analysis included 6 individual results: sodium (SOD), potassium (POTA), chloride (CHLO), bicarbonate (BICAR), urea (URE), creatinine (CREAT).
- The Liver Function Test included 8 individual results: total bilirubin (BILI), conjugated bilirubin (CONJBILI), alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatise (ALK), total protein (PROT), albumin (ALB) and globulin (GLOB).

Figure C2 provides a screenshot showing of the data mask presented on the computer screen to enter data from the laboratory reports.

**Figure C2: Screen capture of the data entry mask**




*Note:* this screenshot shows a (hypothetical) participant taking part in the bonus treatment (the instructions are reminded at the top). The remaining time counter on the top right hand corner indicates that the participant has 449s left; the subject has started working on patient file #8 (see “REF. NUMBER:   8” on the top right-hand corner of the lab report), and so far they have not entered a blood test result yet (shaded areas on the right are still empty). The performance counters on the right indicate that the participant has correctly identified one correct diagnosis, but has not made a single correct data entry.

### Identifying the medical diagnosis

After participants had validated the data entries they had made (they could leave some or all blank if they wanted), another question appeared on the screen that prompted them to make a medical diagnosis based on the patient blood test results (see Figure A3). Based on the test results provided, subjects had then to choose from a proposed list of 13 possible diagnoses: (1) Normal; (2) Normocytic Anaemia; (3) Macrocytic Anaemia; (4) Microcytic Anaemia; (5) Thrombocytopaenia; (6) Pancytopaenia; (7) Leucocytosis; (8) Acute Renal Failure ; (9) Hyponatraemia; (10) Hypokalaemia; (11) Acute Hepatitis; (12) Chronic Hepatitis; (13) Cholestasis

**Figure C3: Screen capture of the diagnosis choice question**


*Note:* this screenshot shows a (hypothetical) participant 203 seconds into the experiment (note that the remaining time counter indicates 397s left): the subject is currently working on patient file #2 (see "REF. NUMBER 2" on the top right-hand corner of the lab report); they entered 8 of the 22 blood test results (all entries are incorrect as the counter in the top right hand corner indicates), then automatically moved on to the diagnosis task: the choice of 13 alternative diagnoses appears on the right hand side of the screen.

As is common practice, the laboratory reports indicated for each result the expected normal reference range, so that it was easy for participants to identify abnormal results. However, participants still had to synthesise the information available from lab report, since not only several results could be abnormal, but the combination of different results could lead to different diagnoses. For example, abnormal results from the Liver Function Test – e.g. alanine aminotransferase (ALT), aspartate aminotransferase (AST) and alkaline phosphatise (ALK) – could lead to a diagnosis of chronic or acute hepatitis depending on other test results.

**Appendix D: Experimental procedures**

We organised a total of 11 sessions. Participants were recruited through adverts posted on their class noticeboard and leaflets distributed at the end of lectures. Adverts mentioned an "original research project", but did not mention any monetary earnings. Participants were invited to pre-book a slot online for one of the sessions, but, provided we had enough space, we also accepted those who showed up on the day even if they had not booked. At the start of a session, participants were randomly given a number corresponding to a workstation, prepared before the session so that it had been randomly allocated to a treatment (in order to avoid confounding the treatment and physical workstation). We sought to achieve an equal number of treatments across sessions in two ways: (1) the numbers given out to participants were chosen from a sequence established to fill workstations from each of the three groups. However in a couple of instances, computer problems forced us to move participants to a nearby workstation, which disrupted the planned ordered allocation; (2) to achieve balance across sessions, some adjustments were made every so often based on the expected number of participants, but we never refused participants to reach perfect balance.

Hence, the final perfect balance of 60 individuals per treatment is partly the result of our process, and partly due to luck (no unexpected participant showed up on the last session).

**Table D1: Participation, by treatment and experimental session**

| **Session No.** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **Total** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Control | 7 | 2 | 6 | 6 | 8 | 1 | 11 | 11 | 3 | 1 | 4 | 60 |
| Gain | 6 | 3 | 6 | 6 | 7 | 2 | 10 | 8 | 5 | 1 | 6 | 60 |
| Loss | 6 | 2 | 7 | 7 | 7 | 1 | 11 | 8 | 3 | 1 | 7 | 60 |
| Total | 19 | 7 | 19 | 19 | 22 | 4 | 32 | 27 | 11 | 3 | 17 | 180 |

## Appendix E: Experiment instructions

Thank you for coming today and for volunteering to take part in this study.
The task to be done today involves entering patient's laboratory results and identifying the abnormal results. There will be two different periods of 10 minutes each. In each period, you will be asked to enter patients' laboratory test results onto the computer and to identify the patients' diagnoses. You will be able to earn money for this.

Each of the laboratory reports that you will be given to enter include a full blood count, a U&E, and a liver function test. Each form has a reference number. This is the number you need to enter when asked for the laboratory reference number. The test results are shown in the shaded blocks. These are the numbers you are required to enter. The report also shows the test units and normal range – you do not enter these.

I will explain the basic aspects of the task and then you will have a training session so that you get familiar with the different aspects of how the data entry masks works. The first screen asks for your study number. The next screen provides some instructions relevant for each period. You will have 90 seconds to read the instructions, make sure you understand them. When you move onto data entry and diagnosis, you will get a screen asking you to enter the reference number of the form. You need to enter the laboratory reference number and press OK to submit this. You will be then be taken to data entry mask and you can start entering data. The numbers you have typed are only entered on the system once you click the RECORD DATA button at the bottom.

Once you clicked the RECORD DATA button the computer will show you a list of possible diagnoses Select which diagnosis you think is correct and press the RECORD DIAGNOSIS button at the bottom to submit your choice and move onto the next record. Once you click a RECORD button you cannot go backwards to change any information.

There are 2 envelopes in front of you, labelled period 1 and period 2. Open the envelope when instructed to do so on the computer screen. In each envelope there are 10 forms provided for you to enter data and diagnose during that period. You can only enter the forms allocated for a particular period, but you can enter the forms in any order you like.
For agreeing to take part in the study, you have already earned R30. But you can also earn additional money for the data entry and diagnosis identification. How you will be paid is explained on the introductory screen at the beginning of each period. You have 90 seconds

to make sure you understand how you will be paid. An example is provided to make the payment system clearer.

In addition, by doing the data entry and identifying the diagnosis, you will also be able to raise money for a charity providing treatments to patients. You will be able to choose your preferred charity from a short list. How much the charity will receive in each period is also explained on the introductory screen.

Your earnings will be paid in cash at the end of the session. We will only pay for 1 of the periods. The period which we will pay will be selected by somebody in the class throwing a die at the end. We will collate the money owing to the various charities and pay them after we have finished the study.

**Screen instructions – Control Treatment**

For entering data and identifying diagnoses, you will be paid R105 for the entire period. This payment does not depend on the number of individual entries or diagnoses you make, or if your entries and diagnoses are correct.

In addition, by completing the data entry and diagnosis identification you will also be able to raise money for a charity of your choice. We will pay your chosen charity R0.20 for each number entered correctly and R1.50 for each diagnosis identified correctly.

**Screen instructions – Gain Treatment**

In this period, for the data entry and diagnosis identification you will be paid R90 for the entire period. This part of the payment does not depend on the number of individual entries or diagnoses you make or if your entries and diagnoses are correct.

In addition, you will be paid an extra bonus for the data items that you enter correctly:

- if you enter between 100 and 109 numbers correctly, you will earn an additional R10
- if you enter between 110 and 119 numbers correctly, you will earn an additional R20
- if you enter between 120 and 129 numbers correctly, you will earn an additional R30
- if you enter between 130 and 139 numbers correctly, you will earn an additional R40
- if you enter more than 140 numbers correctly, you will earn an additional R50

Moreover, by completing the data entry and diagnosis identification you will also be able to raise money for a charity of your choice. We will pay your chosen charity R0.20 for each number entered correctly and R1.50 for each diagnosis identified correctly.

**Screen instructions – Loss Treatment**

In this period, for the data entry and diagnosis identification you will be paid R140 for the entire period. However, if you do not enter enough data correctly, some of this money will be taken away from you: - if you enter more than 140 numbers correctly, you will be paid the full R140

- if you enter between 130 and 139 numbers correctly, you will lose R10
- if you enter between 120 and 129 numbers correctly, you will lose R20
- if you enter between 110 and 119 numbers correctly, you will lose R30
- if you enter between 100 and 109 numbers correctly, you will lose R40
- if you enter less than 100 numbers correctly, you will lose R50.

Finally, by completing the data entry and diagnosis identification you will also be able to raise money for a charity of your choice. We will pay your chosen charity R0.20 for each number entered correctly and R1.50 for each diagnosis identified correctly.

### Appendix F: Correlation between the channels of effort

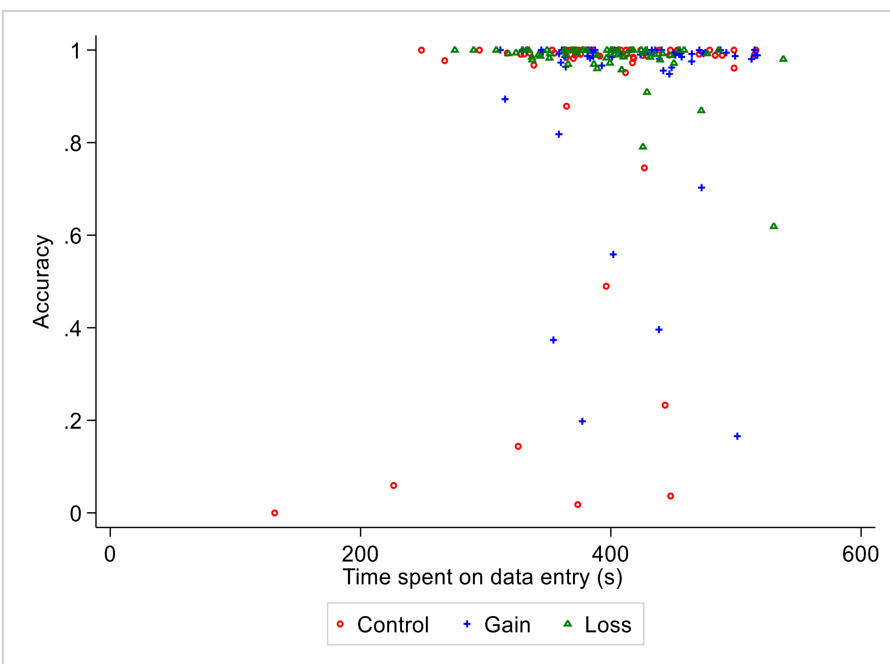
It is possible to imagine that respondents can adopt one of two strategies when trying to increase their performance and minimise errors in data entry:

- *An ex-post* strategy whereby participants seek to minimise errors by double-checking after they have entered a number that it was a correct entry;
- An *ex-ante* strategy where individuals focus their attention before entering numbers, paying attention to each number and avoiding making mistakes.

The first strategy is likely to increase both accuracy and time spent, and lead to a higher correlation between the two measures than the first one. However the second strategy is probably less likely to lead to such correlation: by focusing people's attention, it reduces distraction and makes people be more efficient.

Looking at the data, they do not provide much support to the notion that accuracy and effort are correlated – see the Figure F1 below. The correlation coefficient $\rho = 0.14$ which is quite low. If we exclude the few outliers who did not seem to pay much attention to their work (accuracy<60%), the direction of this correlation is reversed, and still very weak: the more time spent, the lower the accuracy: $\rho = -0.17$. We interpret these low correlations as lending support to the idea that participants avoid mistakes by paying more attention to the task *ex-ante*. Having said that, this should be ideally complemented by an independent measure of cognitive attention (e.g. physiological measure).

**Figure F1: Correlation between effort and accuracy in data entry**



### Appendix G: Analysis of strategic behaviour in the choice of data entries

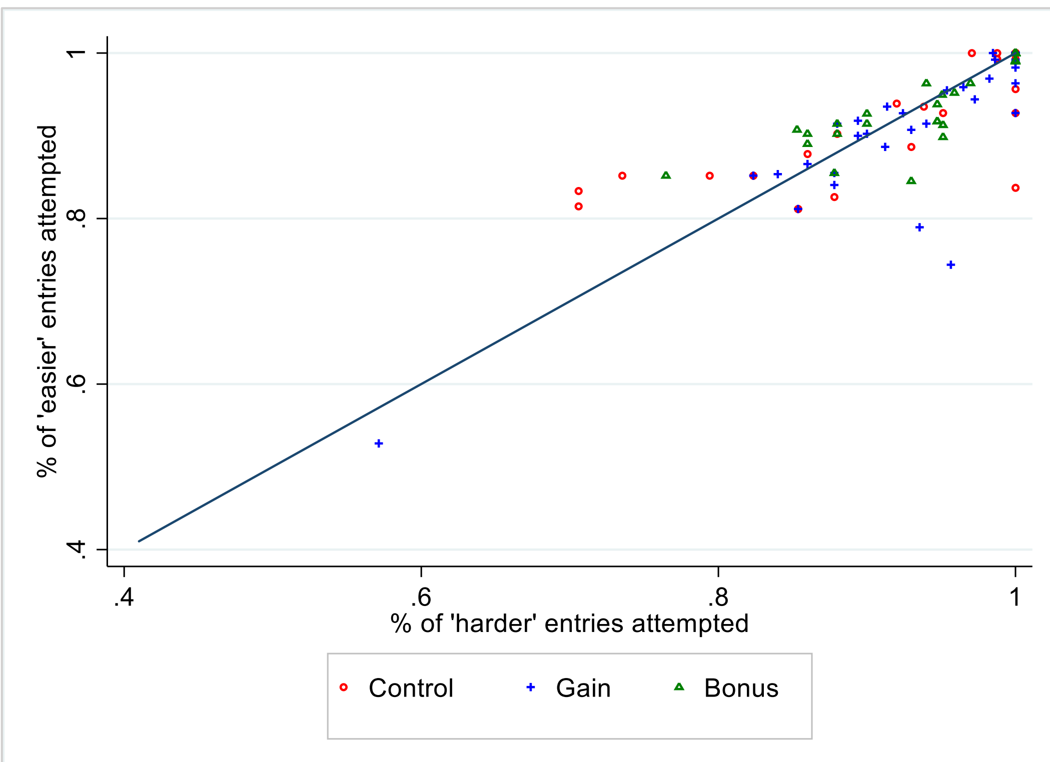
In light of the time constraints they were facing, subjects could have been strategic in the choice of data entries they were attempting, choosing easier ones more than harder ones. We define as an 'easy' entry (blood test result) one that has 1, 2 or 3 characters, which could include numbers and the dot. The following examples would all be considered easy entries: 5 ; 54 ; 5.4 ; 540.

By contrast, a 'hard' data entry includes more than three characters. For example, the numbers 54.4 and 542.1 would both be considered harder.

Overall, the ten patient files given to participants in the period analysed in the paper included 9 one-character entries; 18 two-character entries; 54 three-character entries; 113 four-character entries and 26 five-character entries.

Figure G1 below shows the relation proportion of easy and hard entries attempted by participants and does not suggest any particular strategic behaviour.

**Figure G1: Choice of easier vs. harder blood test results attempted**




Notes: 'Easier' entries are those that include 3 or fewer characters (including digit(s) and a dot if needed); 'harder' entries require 4 or 5 characters (digits and a dot if needed). The 45-degree line indicates that respondents attempted hard and easy entries in an equal proportion.

## Appendix H: Optimal strategy in effort allocation

The choice to focus on data entry or diagnosis depends on individuals' expected earnings in each activity, which are determined by: the social reward associated with each correct output ($w_i$); the total number of outputs done in a period (correct or not), which itself depends on the average time per output produced ($t_i$); and the accuracy rate $acc_i$ in activity *i*.

A rational decision-maker should focus on data entry as long as their expected earnings from data entry are greater than those from the diagnosis activity:

$$R0.20 \times \frac{600}{t_{de}} \times acc_{de} > R1.50 \times \frac{600}{t_{dia}} \times acc_{dia} \qquad (1)$$

We can simplify each term of this inequality by using the average time per correct answer $tc_i$ defined as:

$tc_i = \frac{600}{Nc_i}$ where $NC_i$ is the total number of correct outputs for activity *i*.

Since $t_i = \frac{600}{N_i}$ (where $N_i$ is the total number of attempts for activity *i*), and $acc_i = \frac{Nc_i}{N_i}$ , we can rearrange $\frac{600}{t_i} \times acc_i$ as $600 \times \frac{N_{de}}{600} \times \frac{NC_{de}}{N_{de}} = 600 \times \frac{1}{tc_{de}}$.

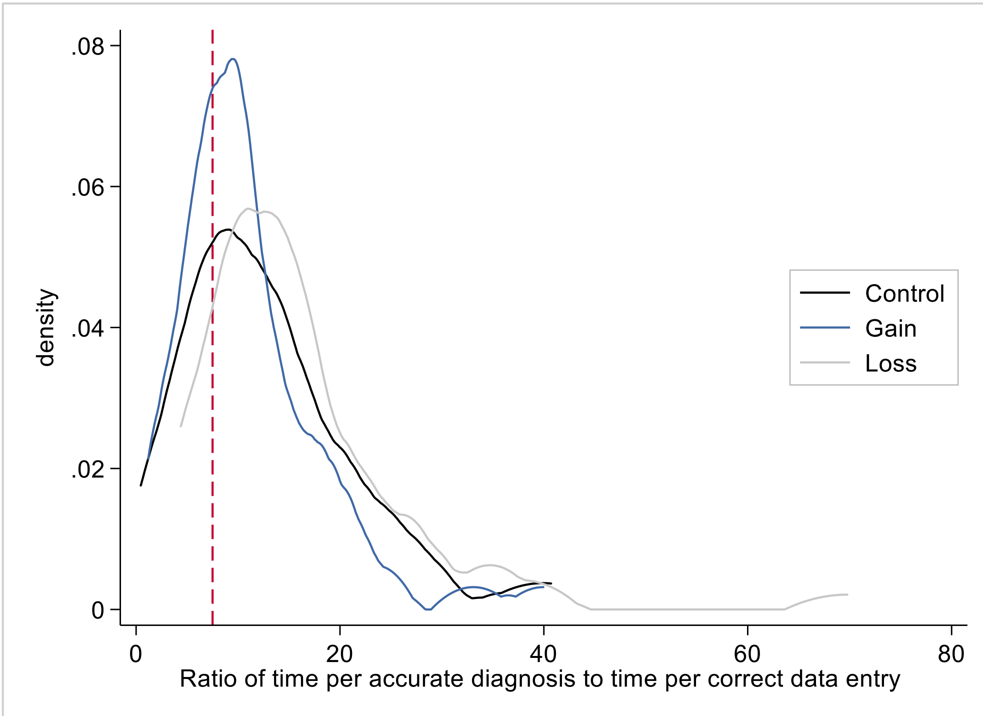
Hence (1) can be re-written as:

$$tc_{dia} > 7.5\ tc_{de} \qquad (2)$$

In other words, a subject should focus on data entry if the time s-he needs to obtain a *correct* diagnosis is at least 7.5 times higher than the time per correct data entry.

Figure H1 below presents the distribution of the ratio $\frac{tc_{dia}}{tc_{de}}$ in our sample, and clearly shows that even if they seek to maximise social benefits, it is optimal for most individuals to focus on data entry. Focusing on the diagnosis task is only optimal for 16.7% of respondents on the loss treatment, and 28.3% in the gain and control treatments.

**Figure H1: Distribution of the relative abilities in the two activities**




*Note*: the dashed vertical line indicates the ratio of relative abilities of diagnosis to data entry beyond which it is optimal for participants to focus on data entry to maximise social benefits

In practice, the actual choice that individuals face is slightly more complex than the simplified framework above, because their number of attempts in both activities is limited to 10 patient files in one period, which translates into a maximum of 10 diagnoses and 220 data entries attempted. In other words, even if a subject has a clear comparative advantage in one activity, they may not be able to spend *all* of their time on it: if they are efficient enough in that activity, they will need less than 600 seconds to make all their attempts. Although a subject cannot go back to a case already submitted, we could imagine that subjects adjust their effort allocation decisions over the course of the period. In the illustration below, we assume that any spare time left after finishing all cases in one activity could have been re-allocated seamlessly to the other activity.

Table H1 below illustrates the outcomes of extreme strategies where individuals prioritize one activity over the other. We assume that it takes about 11 seconds to register a new patient – that is the median registration time per case observed in the control group. Next, we determine the time needed and performance achieved for a given case, depending on subject's abilities (depicted in Panel A).

**Table H1: Outcomes of strategies prioritizing one activity**

| | Median abilities in data entry and diagnosis | Median abilities in data entry, top abilities in diagnosis | Top abilities in data entry and diagnosis |
|---|---|---|---|
| **PANEL A: ASSUMPTIONS** | (A1) | (A2) | (A3) |
| Registration time per patient (s) | 10.96 | 10.96 | 10.96 |
| **Data entry abilities** | | | |
| Accuracy | 0.99 | 0.99 | 1.00 |
| Av. time per correct entry | 4.12 | 4.12 | 2.57 |
| Av. # correct entries per sheet | 21.78 | 21.78 | 22.00 |
| Time per sheet (incl. registration) | 100.69 | 100.69 | 67.50 |
| **Diagnosis abilities** | | | |
| Accuracy | 0.73 | 1.00 | 1.00 |
| Av. time per correct diagnosis | 47.95 | 22.38 | 22.38 |
| Average # diagnosis per sheet | 0.73 | 1.00 | 1.00 |
| Time per case (incl. registration) | 45.96 | 33.34 | 33.34 |
| **Ratio of abilities** | 11.64 | 5.43 | 8.71 |

| **PANEL B: PERFORMANCE** | If prioritize data entry (B1) | If prioritize diagnosis (B2) | If prioritize data entry (B3) | If prioritize diagnosis (B4) | If prioritize data entry (B5) | If prioritize diagnosis (B6) |
|---|---|---|---|---|---|---|
| Number of patient cases registered | 6 | 10 | 6 | 10 | 9 | 10 |
| **Data entry** | | | | | | |
| Time spent on data entry | 534.24 | 140.37 | 534.24 | 266.60 | 501.36 | 266.60 |
| Total # of correct entries | 129.67 | 34.07 | 129.67 | 64.71 | 195.08 | 103.74 |
| Total social earnings | 25.93 | 6.81 | 25.93 | 12.94 | 39.02 | 20.75 |
| **Diagnosis** | | | | | | |
| Time on diagnosis | 0 | 350.04 | 0 | 223.80 | 0 | 223.80 |
| Total # of correct diagnoses | 0 | 7.30 | 0 | 10.00 | 0 | 10.00 |
| Total social earnings | 0 | 10.95 | 0 | 15.00 | 0 | 15.00 |
| **Total social earnings** | 25.93 | 17.76 | 25.93 | 27.94 | 39.02 | 35.75 |

*Notes:* Median abilities are defined based on the median performance in the control group; "top" abilities correspond to the performance of the top decile in the control group. The ratio of abilities is defined as $\frac{tc_{dia}}{tc_{de}}$ above, or the ratio of the average time per correct diagnosis over the average time per correct data entry. In Panel B, the performance shown in columns B1 and B2 are based on assumptions made for an individual described in Panel A, column A1. Similarly, the performance shown in columns B3 and B4 (B5 and B6) are based on assumptions made for an individual described in Panel A, column A2 (A3).

For a subject with median abilities in both activities (column A1) – i.e. with a median accuracy of 0.99 in data entry and an average time per correct entry of 4.12 seconds – it would take 100.69 seconds to do a whole sheet (22 entries), including the time for registration; this would result in 21.78 correct entries on average per sheet. It follows that this median individual could do 6 cases (5 full cases, and one partially completed) and 129.67 correct entries, spending a total of 523.24 seconds on data entry (the rest of patient registration). If the same individual prioritised diagnosis (column B2), they would end up with 7.3 correct diagnoses and some time left to enter data, but overall their performance would be lower and they would raise less money.

For a subject with median abilities in data entry but top abilities in diagnosis (column A2), prioritising the diagnosis activity would actually be the optimal solution. This individual would be able to diagnose all ten cases correctly (based on the top accuracy rate in diagnosis of 100%), and since they would do this relatively quickly, they could also spend 266.6 seconds on data entry and make 64.7 correct entries. Overall, they would raise about R2 more than if they had only focused on data entry (column B3, which is identical to column B1).

Finally, for a subject with top abilities in data entry *and* diagnosis (column A3), prioritising the diagnosis activity is no longer the optimal solution because the time spent on data entry has a higher return. If they prioritized the diagnosis identification (column B6), they would diagnose all ten cases correctly and make an additional 103.7 correct entries. This would allow them to raise R35.75, falling short by about R3 of what they would raise by focusing on data entry only (see column B5).