

# BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email [editorial.bmjopen@bmj.com](mailto:editorial.bmjopen@bmj.com)

# BMJ Open

## Clinical decision support using Labtracker+: a medical app for the interpretation of consecutive laboratory results

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2017-015854
Article Type:	Research
Date Submitted by the Author:	06-Jan-2017
Complete List of Authors:	Hilderink, Judith; Maastricht University Medical Center+, Central Diagnostic Laboratory, Clinical Chemistry Rennenberg, Roger; Maastricht University Medical Center+, Internal Medicine Vanmolkot, Floris; Maastricht University Medical Center+, Internal Medicine Bekers, Otto; Maastricht University Medical Center+, Central Diagnostic Laboratory, Clinical Chemistry Koopmans, Richard; Maastricht University Medical Center+, Internal Medicine Meex, Steven; Maastricht University Medical Center+, Central Diagnostic Laboratory, Clinical Chemistry
<b>Primary Subject Heading</b>:	Medical education and training
Secondary Subject Heading:	Diagnostics
Keywords:	Medical application, Biological variation, MEDICAL EDUCATION & TRAINING

SCHOLARONE™  
Manuscripts

only

**Full title**

Clinical decision support using *Labtracker+*: a medical app for the interpretation of consecutive laboratory results

**List of authors**

J.M. Hilderink<sup>1</sup>, [judith.hilderink@mumc.nl](mailto:judith.hilderink@mumc.nl)

R.J.M.W. Rennenberg<sup>2</sup>, [r.rennenberg@mumc.nl](mailto:r.rennenberg@mumc.nl)

F.H.M. Vanmolkot<sup>2</sup>, [floris.vanmolkot@mumc.nl](mailto:floris.vanmolkot@mumc.nl)

O. Bekers<sup>1</sup>, [o.bekers@mumc.nl](mailto:o.bekers@mumc.nl)

R.P. Koopmans<sup>2</sup>, [richard.koopmans@mumc.nl](mailto:richard.koopmans@mumc.nl)

S.J.R. Meex<sup>1</sup>, [steven.meex@mumc.nl](mailto:steven.meex@mumc.nl)

**Affiliations**

<sup>1</sup> Department of Clinical Chemistry, Central Diagnostic Laboratory, Maastricht University Medical Centre+, Maastricht, the Netherlands

<sup>2</sup> Department of Internal Medicine, Maastricht University Medical Centre+, Maastricht, the Netherlands

**Corresponding author**

Steven J. R. Meex

Department of Clinical Chemistry, Cardiovascular Research Institute Maastricht (CARIM),

1  
2  
3 Maastricht University Medical Centre+ (MUMC+)

4  
5  
6 P.O. Box 5800, 6202 AZ Maastricht, the Netherlands

7  
8  
9 Telephone number: 31-(0)43-3874694

10  
11 Fax number: 31-(0)43-3874692

12  
13 [14 steven.meex@mumc.nl](mailto:steven.meex@mumc.nl)

15  
16  
17  
18  
19 **Word count main text**

20  
21 2201

22  
23  
24  
25  
26 **Keywords**

27  
28 Medical application, biological variation, medical education

## ABSTRACT

**Objectives:** When monitoring patients over time, clinicians may struggle to distinguish ‘true changes’ in consecutive blood parameters from so-called natural fluctuations. In practice, they have to do so by relying on their clinical experience and intuition. We developed *Labtracker+*, a medical app that calculates the probability that an increase or decrease over time in a specific blood parameter is true, given the time between measurements.

**Design:** We presented patient cases to 135 participants to examine whether there is a difference between medical students, residents and experienced clinicians when it comes to interpreting changes between consecutive laboratory results. Participants were asked to interpret if changes in consecutive laboratory values were likely to be ‘true’ or rather due to natural fluctuations. The answers of the study participants were compared to the calculated probabilities by the app *Labtracker+* and the concordance rates were assessed.

**Setting and participants:** Medical students ( $n=92$ ), medical residents from the department of internal medicine ( $n=19$ ) and internists ( $n=24$ ) at a Dutch university medical centre.

**Primary and secondary outcome measures:** Concordance rates between the study participants and the calculated probabilities by the app *Labtracker+* were compared. Besides, we tested whether physicians with clinical experience scored better concordance rates with the app *Labtracker+* than inexperienced clinicians.

**Results:** Medical residents and internists showed significantly better concordance rates with the calculated probabilities by the app *Labtracker+* than medical students, regarding their interpretation of differences between consecutive laboratory results ( $p=0.009$  and  $p<0.001$ , respectively).

1  
2  
3 **Conclusion:** The app *Labtracker+* could serve as a clinical decision tool in the interpretation of  
4 consecutive laboratory test results, especially for inexperienced clinicians.  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

## STRENGTHS AND LIMITATIONS OF THIS STUDY

- This study describes an objective scientific approach of distinguishing a ‘true change’ from a ‘random fluctuation’, when interpreting a change between two consecutive laboratory measurements
- The database that was used to develop the app Labtracker+ presently contains only a limited number of laboratory parameters

For peer review only

## INTRODUCTION

Medical doctors frequently monitor the laboratory results of patients to determine whether their condition is stable, improving or deteriorating, and to decide whether treatment should be continued, initiated, or rather be postponed. Hence, an intuitive ‘reference frame’ of the patient can be formed by repeated assessment of laboratory parameters. An expertise-based reference frame is often lacking in less experienced clinicians, such as medical residents, or medical students. Hence, it can be difficult for less experienced clinicians to distinguish natural fluctuations in laboratory results from ‘true changes’ and vice versa.

According to a study of Jones et al., 1 in 5 medical graduates identified themselves as being ‘less than competent’ in using laboratory testing.<sup>1</sup> Additionally, a study of Hickner et al. demonstrated that even the experienced clinicians reported uncertainty in interpreting laboratory test results in approximately 10% of the ordered laboratory tests.<sup>2</sup>

To help (inexperienced) clinicians to make the distinction whether a change between measurements is true or not, we developed a medical app (*Labtracker+*), that calculates the probability of a ‘true change’ between two consecutive laboratory results. It is simple to use in clinical practice and provides evidence-based decision support to complement intuitive interpretation of changes in consecutive laboratory results.

The current study tests the hypothesis that experienced clinicians show better concordance rates with *Labtracker+*-based calculations than medical students and less experienced clinicians.



## METHODS

This study was approved by the Dutch Association for Medical Education (NVMO). Study participants provided written informed consent.

### Study participants

135 participants were included: 92 medical students, 19 medical residents, and 24 medical specialists. We scheduled six days on which medical students participated in our study. All students were at the same stage of their internship and 92% of them were in their first medical internship (usually in their fourth year of the six years training to become a medical doctor in the Netherlands). The medical residents and medical specialists, all employed at the department of Internal Medicine at Maastricht University Medical Centre, were recruited during the daily morning briefing. The mean number of clinical experience years was  $24 \pm 9$  for the internist and  $3.5 \pm 2$  for the residents.

Exclusion criteria were: medical students in the bachelor phase of their studies and residents or specialists of a specialty other than internal medicine.

### Study design

Ten fictive clinical cases were presented to the study participants. These cases were conceived and developed by the authors of this article. Central to each case was to interpret the likelihood of change between consecutive measurements of a certain laboratory parameter. The time interval between the measurements varied for all cases. All cases can be found in the supplemental data.

1  
2  
3 The participants were asked to categorize the presented changes in one of four likelihoods:  
4  
5 ‘unlikely’ (<50%), ‘doubtful’ (50-80%), ‘probably’ (80-95%), and ‘very likely’ (> 95%).  
6  
7

8 Participants answered these questions individually. The extent to which the responses of the  
9  
10 participants were in accordance with the calculated probabilities by the app *Labtracker+* were  
11  
12 analysed.  
13

14  
15 The laboratory parameters that were incorporated in the cases were: haemoglobin, leukocytes,  
16  
17 thyroid stimulating hormone (TSH), vitamin B12, alanine aminotransferase (ALAT), glucose,  
18  
19 glycosylated haemoglobin type A1c (HbA1c), creatinine, activated partial thromboplastin time  
20  
21 (aPTT) and C-reactive protein (CRP).  
22  
23

### 24 25 26 27 28 **Calculating the probability of change using the app *Labtracker+***

29  
30 For a change between consecutive measurements to become significant, the difference must be  
31  
32 larger than the change that would reasonably be expected due to normal analytical- and  
33  
34 biological variation. This is termed the reference change value (RCV) and was first described by  
35  
36 Fraser et al.<sup>3</sup> The RCV has become an established concept in laboratory medicine.  
37

38  
39 The RCV differs for each parameter and depends on the within-person biological variation ( $CV_I$ )  
40  
41 and the analytical variation ( $CV_A$ ) and can be calculated as follows:<sup>4-5</sup>  
42  
43

$$44  
45  
46 RCV = \sqrt{2} * Zscore * \sqrt{CV_A^2 + CV_I^2}$$

47  
48  
49 In this formula, the Z-score represents the number of standard deviations and corresponds to the  
50  
51 desired probability. Commonly used Z-scores are 1.96 and 2.56. These Z-scores calculate the  
52  
53 percentage increase or decrease that is required to become statistically significant, with a false  
54  
55 positive rate of 5%, ( $p < 0.05$ ) and 1% ( $p < 0.01$ ), respectively.  
56  
57  
58  
59  
60

1  
2  
3 An important side note to calculate the RCV is that not all clinical decisions are made with  
4 statistical probabilities of >95%. In other words, some changes are strictly not statistically  
5 significant, but the probability may be high enough to lead to a treatment or clinical decision. For  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

An important side note to calculate the RCV is that not all clinical decisions are made with statistical probabilities of >95%. In other words, some changes are strictly not statistically significant, but the probability may be high enough to lead to a treatment or clinical decision. For example, a haemoglobin concentration of 6.1 mmol/L was measured in a male patient. After a week, it dropped to 5.8 mmol/L. Although the likelihood probability of a 'real' decrease in haemoglobin level is less than 95% in this case - namely around 70% - the clinician may still decide to use additional diagnostic testing or to treat the patient.

For implementation in the smartphone app, we rearranged the RCV-formula so that the Z-score (and therefore the likelihood probability) is the unknown. The probability of a real change between consecutive measurements will then be calculated. The rearranged RCV-formula is:

$$Z_{\text{score}} = \frac{\text{percentage change between consecutive measurements}}{\sqrt{2 * (CV_A^2 + CV_I^2)}}$$

### Statistical analysis

All answers given by the study participants were compared to the calculated probabilities by the app Labtracker+. Two statistical analyses were performed. Firstly, differences among the study groups were calculated using one-way analysis of variance (ANOVA). *p*-values < 0.05 were considered statistically significant. Comparisons between pairs of groups were adjusted for multiple testing using the Bonferroni correction.

Secondly, two groups were compared; the medical students formed the 'inexperienced group' and the medical residents and internists together formed the 'experienced group'. The two groups were compared by using the Fisher's exact test. All statistical calculations were performed by using SPSS version 23 (IBM SPSS Statistics. IBM Corporation, Armonk, New York).

## RESULTS

Data collection was incomplete for two subjects (1.5%); it concerns two medical students who did not fill in all questions and therefore, they were excluded from all analyses.

### Psychometric analysis

We performed the 'Item Test Correlation ( $R_{IT}$ )' test, a psychometric analysis to check the distinctiveness of the case questions.<sup>6 7</sup> Briefly, the  $R_{IT}$  is a correlation coefficient which can vary between -1 and 1; a question with a  $R_{IT}$  score of 0 does not distinguish between good and poor performers. If the  $R_{IT}$  score is negative then the question may not be valid and should be removed.<sup>6 7</sup> The  $R_{IT}$  score of the aPTT-question had a negative value and therefore, this question was excluded from further analyses. The  $R_{IT}$  score of the other questions were all considered sufficient.

### Comparison of the app *Labtracker+* with the estimates of physicians

The correct answer was defined as the answer that matched exactly with the calculated probability by the app *Labtracker+*, i.e. the answer category containing the calculated probability percentage.

In figure 1, the estimates of the study groups were shown. Participants with extensive clinical experience showed higher concordance rates with *Labtracker+*- calculations; with internists and the medical residents scoring significantly better than the medical students ( $p < 0.001$  and  $p = 0.021$ , respectively).

The percentage of questions answered correctly by the medical residents did not differ significantly from the internists ( $p = 0.719$ ; 95% CI -5.4 – 15.5).

1  
2  
3 A sensitivity analysis was conducted that allowed a broader range to define concordant answers.  
4  
5 In this sensitivity analysis, not only the answer category that matched exactly with the calculated  
6  
7 probability by *Labtracker*<sup>+</sup> was considered correct, but also the answer category adjacent to that  
8  
9 category. The results of this analysis are shown in figure 2. On average, the medical students  
10  
11 answered 69% of the questions correctly, the medical residents 78% and the internists 81%.  
12  
13 Similar to the primary analysis, concordance rates of answers from internists with the  
14  
15 *Labtracker*-calculated probabilities were significantly better than medical students ( $p<0.001$ ;  
16  
17 95% CI 4.7 – 19.0). Again, the difference between medical residents and medical students was  
18  
19 statistically significant as well ( $p=0.020$ ; 95% CI 1.1 – 16.8). And, like the first analysis,  
20  
21 concordance rates between the medical residents and *Labtracker*<sup>+</sup> did not differ significantly  
22  
23 from the internists ( $p=1.000$ ; 95% CI -12.5 – 6.6). Therefore, in subsequent analyses, internists  
24  
25 and medical residents were pooled as one group (the experienced clinicians) and compared with  
26  
27 medical students.  
28  
29  
30  
31  
32  
33  
34  
35  
36

### 37 **Analysis per parameter**

38  
39 The scores of the study participants for each parameter were assessed separately. Figure 3  
40  
41 depicts the percentage of correct answers from the study groups per parameter. The  
42  
43 ‘experienced’ group (the residents and the internists) was compared with the medical students. In  
44  
45 general, the experienced group demonstrated the most concordant answers, compared to the  
46  
47 inexperienced medical students. The experienced clinicians scored best on the parameter  
48  
49 haemoglobin (95%) and the parameter they scored the worst on was creatinine (28%). The  
50  
51 medical students scored best on the haemoglobin parameter as well (82%). They scored worst on  
52  
53 the TSH question (6%).  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 In accordance with the previous analysis, we conducted a sensitivity analysis in which the  
4 category adjacent to the correct answer was also considered to be correct. These results are  
5 shown in figure 4. Again, the experienced clinician gave the most concordant answers. In this  
6  
7  
8  
9  
10 analysis, the experienced group (residents and internists) scored 88% correct answers on average.  
11  
12 They scored the best on the vitamin B12, leukocytes, and haemoglobin questions (98%). The  
13  
14  
15 TSH question was the least well answered of all parameters (70%). The medical students scored  
16  
17  
18 90% correct answers on the haemoglobin question. The least concordant question by the students  
19  
20 was TSH (44%).  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## DISCUSSION

The current study demonstrated that experienced clinicians show better concordance rates with the calculated probabilities by the app *Labtracker+* than less experienced medical students.

Stratified analyses reveal that this pattern holds true for the majority of laboratory parameters studied. Concordance rates with the app demonstrate an incremental improvement from medical students, to residents and internists. The differences between clinicians with experience (internists and residents) and inexperienced medical students were statistically significant. This outcome illustrates the relevance and added value of the medical app *Labtracker+*.

The clear association between the degree of clinical experience and improved concordance rates with the calculated probabilities by the app supports the validity of *Labtracker+* calculations.

Improved concordance rates with increasing clinical experience suggest that *Labtracker+* may well be of benefit for less experienced clinicians, such as medical students.

The benefit offered by the app may be parameter dependent. The concordance rates with the calculated probabilities varied from 6% for students that answered the TSH question to 98% for the haemoglobin question for experienced clinicians. This difference in concordance rates suggests that the ‘number intuition’ of a clinician differs per laboratory parameter and varies with clinical experience. It may be difficult for a clinician to estimate how a ‘normal value’ of a parameter fluctuates over time in an individual if the clinician has little actual experience with that parameter. In other words, a clinician with more experience in interpreting a specific parameter can better estimate its biological variation. Since the scores of the medical residents were not significantly different from those of the internists, it might be assumed that medical

1  
2  
3 residents develop their ‘number intuition’ early in the residency training.  
4

5 Not all parameters were scored equally concordant by the participants. A potential explanation  
6 for the ‘worse scored questions’ could be that not all laboratory parameters are widely used in  
7 clinical practice, and that clinicians therefore have little experience in interpreting those  
8 parameters. Another potential explanation could be that not all questions were of equal difficulty.  
9  
10  
11  
12  
13

14  
15  
16  
17 Some limitations in the current study merit attention. Firstly, the groups of medical residents and  
18 specialists were relatively small ( $n=19$  and  $n=24$ , respectively). Secondly, only nine parameters  
19 were tested, with one single interpretation per parameter. However, our estimation is that the  
20 study groups are a good reflection of the total population of people in these groups. In addition,  
21 the parameters we studied are commonly used in clinical practice and cover a broad range of  
22 clinical domains. Therefore, it can reasonably be expected that the results of this study can be  
23 generalized to all commonly used laboratory parameters.  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33

34 Thirdly, it remains unclear whether the use of the medical app *Labtracker+*, would affect clinical  
35 decision making. In that respect, a study that incorporated medical decision making, based on  
36 hypothetical cases involving laboratory test result interpretation(s), would be an interesting  
37 addition. The hypothesis that could then be tested is whether inexperienced clinicians make the  
38 same decisions as more experienced clinicians when they use the app *Labtracker+*.  
39  
40  
41  
42  
43  
44  
45  
46  
47

48 In 2014, Abbot et al. demonstrated that the self-assessed knowledge about interpreting laboratory  
49 results of inexperienced medical residents significantly improved after receiving an educational  
50 program about ‘the basics of laboratory test ordering and interpretation.’<sup>8</sup> This finding endorses  
51 the importance of education for inexperienced clinicians. Elaborating on this, it would be  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 interesting to investigate whether *Labtracker+*, as an educative aid, could provide better medical  
4  
5 decision making.  
6  
7  
8  
9

10 In conclusion, the current study has provided insight into the intuition of medical students with  
11 hardly any clinical experience, clinicians with some experience (medical residents from the  
12 department of internal medicine), and experienced clinicians (internists), when interpreting a  
13 difference between consecutive laboratory results. The findings suggest the existence of a trend  
14 that the ‘number intuition’ of a clinician improves with more years of experience. The  
15 development of this intuition fosters a better interpretation of laboratory results and could  
16 possible improve clinical decision making. In that sense, the use of the medical app *Labtracker+*  
17 can be a useful tool, especially for inexperienced clinicians.  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## REFERENCES

1. Jones A, McArdle PJ, O'Neill PA. Perceptions of how well graduates are prepared for the role of pre-registration house officer: a comparison of outcomes from a traditional and an integrated PBL curriculum. *Med Educ* 2002;**36**(1):16-25.
2. Hickner J, Thompson PJ, Wilkinson T, et al. Primary care physicians' challenges in ordering clinical laboratory tests and interpreting results. *J Am Board Fam Med* 2014;**27**(2):268-74.
3. Fraser CG. Reference change values. *Clin Chem Lab Med* 2012;**50**(5):807-12.
4. Ricos C, Alvarez V, Perich C, et al. Rationale for using data on biological variation. *Clin Chem Lab Med* 2015;**53**(6):863-70.
5. Ricos C, Alvarez V, Cava F, et al. Current databases on biological variation: pros, cons and progress. *Scand J Clin Lab Invest* 1999;**59**(7):491-500.
6. Ebel RL, Frisbie DA. *Essentials of Educational Measurement*. Pearson, 5th edition 1991.
7. Ory JC, Ryan KE. *Tips for Improving Testing and Grading*. Sage Publications Inc 1993.
8. Abbott M, Paulin H, Sidhu D, et al. Laboratory tests, interpretation, and use of resources: a program to introduce the basics. *Can Fam Physician* 2014;**60**(3):e167-72.

## AUTHOR'S CONTRIBUTION

Concept and design of the study, interpretation of the data and drafting of the manuscript: JMH, SJRM and RPK. Statistical analysis: JMH. Critically revising the manuscript and providing important intellectual content: RJMWR, FHMV, OB.

## FUNDING

This work was supported by an Academic Incentive Research Grant from Maastricht University Medical Center+ to dr. Meex.

## COMPETING INTERESTS

*Labtracker+* was conceived and developed by J. Hilderink, R. Koopmans, R. Rennenberg, and S. Meex. It is a free medical app, and therefore financial interests do not apply.

## FIGURE CAPTIONS

### Figure 1

Title: Percentage concordant answers per group

Caption: Values are mean percentages  $\pm$  SD.

Medical students vs. internists,  $p < 0.001$ ; 95% CI 6.96–22.59, medical students vs. medical residents,  $p = 0.009$ ; 95% CI 1.10–18.28, medical residents vs. internists,  $p = 0.321$ ; 95% CI -5.36–15.53.

$p$ -values  $< 0.05$  were considered statistically significant (\*).

### Figure 2

Title: Percentage concordant answers when also the probability adjacent to the designated category was considered correct.

Caption: Values are mean percentages  $\pm$  SD.

Medical students vs. internists,  $p = 0.007$ ; 95% CI 4.72–19.00, medical students vs. medical residents,  $p = 0.031$ ; 95% CI 1.07–16.76, medical residents vs. internists,  $p = 0.609$ ; 95% CI -5.36–15.53.

$p$ -values  $< 0.05$  were considered statistically significant (\*).

### Figure 3

Title: Percentage concordant answers per parameter

Caption: TSH  $p < 0.001$ , creatinine  $p = 0.012$ , vitamin B12  $p = 0.024$ , HbA1c  $p = 0.036$ , haemoglobin

1  
2  
3  $p=0.056$ , ALAT  $p=0.237$ , leukocytes  $p=0.307$ , glucose  $p=0.462$ , CRP  $p=0.574$

4  
5  
6 In this figure, the medical residents and the internists were combined into the group 'clinicians  
7  
8 with experience'  
9

10  
11  
12  
13  
14 **Figure 4:**

15  
16 Title: Percentage concordant answers per parameter when also the probability adjacent to the  
17  
18 designated category was considered correct.  
19

20  
21 Caption: Vitamin B12  $p=0.002$ , TSH  $p=0.006$ , leukocytes  $p=0.012$ , haemoglobin  $p=0.167$ ,  
22  
23 creatinine  $p=0.012$ , ALAT  $p=0.221$ , glucose  $p=0.329$ , HbA1c  $p=1.000$ , CRP  $p=1.000$ .  
24

25  
26 In this figure, the medical residents and the internists were combined into the group 'clinicians  
27  
28 with experience'  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

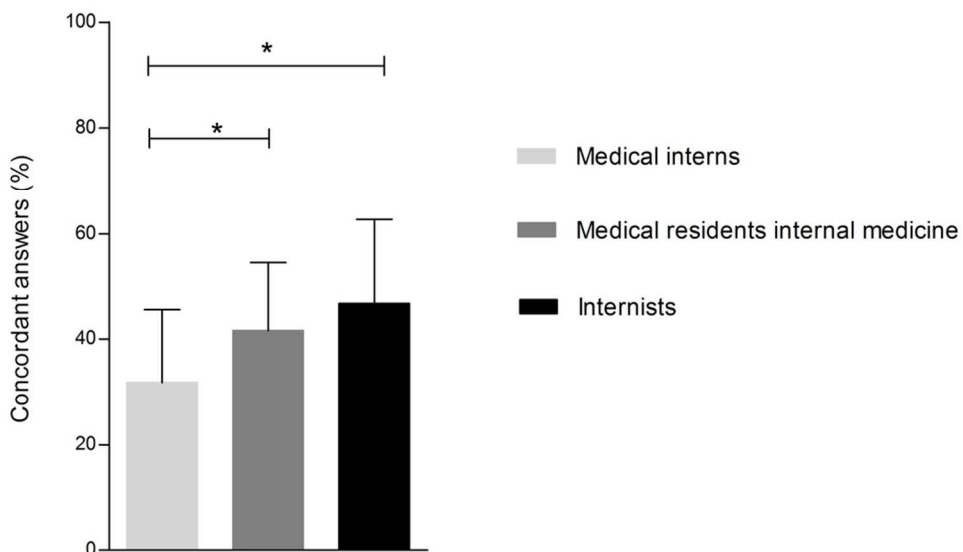


Figure 1. Percentage concordant answers per group

Values are mean percentages  $\pm$  SD.

Medical students vs. internists,  $p < 0.001$ ; 95%CI 6.96–22.59, medical students vs. medical residents,  $p = 0.009$ ; 95% CI 1.10–18.28, medical residents vs. internists,  $p = 0.321$ ; 95% CI -5.36–15.53.  $p$ -values  $< 0.05$  were considered statistically significant (\*).

92x57mm (300 x 300 DPI)

ew only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

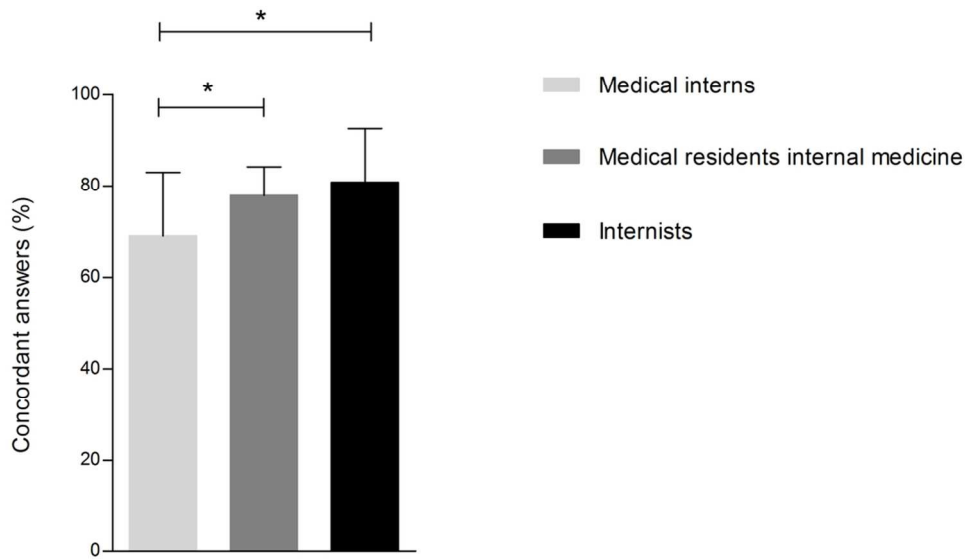


Figure 2. Percentage concordant answers when also the probability adjacent to the designated category was considered correct

Values are mean percentages ± SD.  
Medical students vs. internists, p=0.007; 95% CI 4.72–19.00, medical students vs. medical residents, p=0.031; 95% CI 1.07–16.76, medical residents vs. internists, p=0.609; 95% CI -5.36–15.53.  
p-values < 0.05 were considered statistically significant (\*)

96x60mm (300 x 300 DPI)

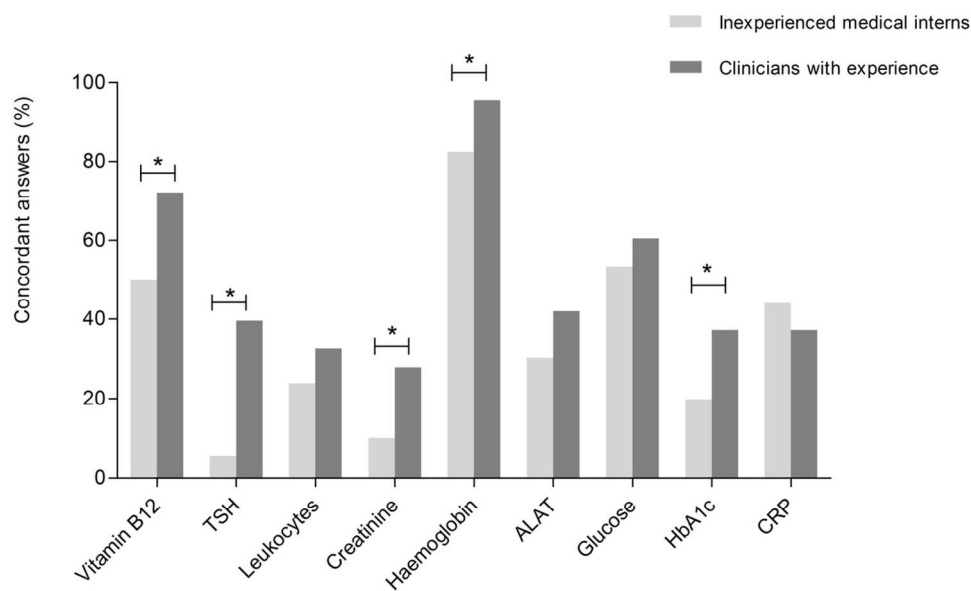


Figure 3. Percentage concordant answers per parameter

TSH  $p < 0.001$ , creatinine  $p = 0.012$ , vitamin B12  $p = 0.024$ , HbA1c  $p = 0.036$ , haemoglobin  $p = 0.056$ , ALAT  $p = 0.237$ , leukocytes  $p = 0.307$ , glucose  $p = 0.462$ , CRP  $p = 0.574$

In this figure, the medical residents and the internists were combined into the group 'clinicians with experience'

107x66mm (300 x 300 DPI)

Peer Review Only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



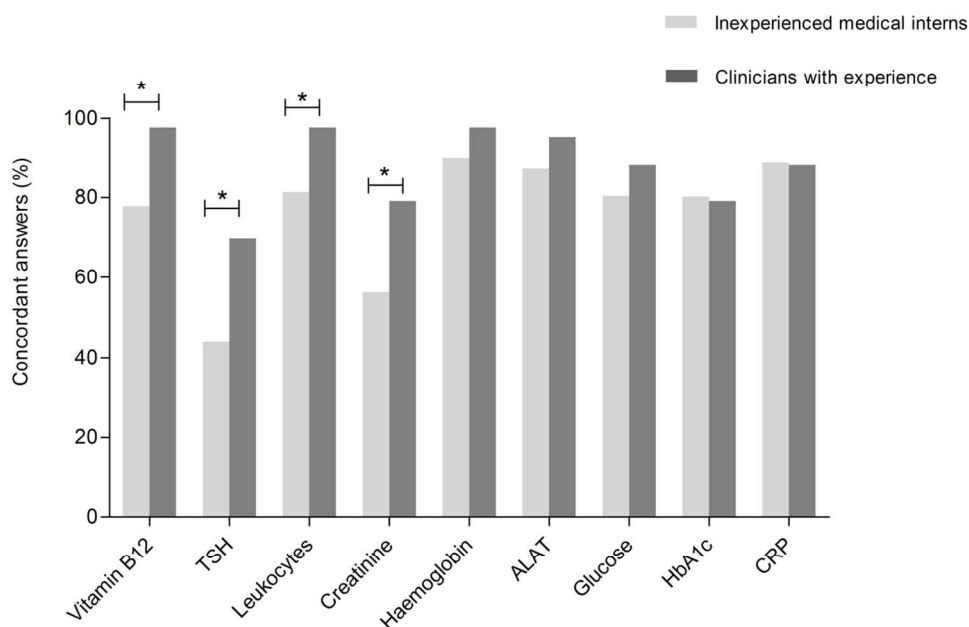


Figure 4. Percentage concordant answers per parameter when also the probability adjacent to the designated category was considered correct

Vitamin B12  $p=0.002$ , TSH  $p=0.006$ , leukocytes  $p=0.012$ , haemoglobin  $p=0.167$ , creatinine  $p=0.012$ , ALAT  $p=0.221$ , glucose  $p=0.329$ , HbA1c  $p=1.000$ , CRP  $p=1.000$ .

In this figure, the medical residents and the internists were combined into the group 'clinicians with experience'

113x75mm (300 x 300 DPI)

## Supplementary material: Case questions

Correct answers, calculated by the app *Labtracker+*, are shown in bold.

Reference values are from Maastricht University Medical Center (MUMC+).

### 1. Glucose

**Reference values fasting glucose [4.0 – 7.0 mmol/L]**

First measurement: **[7.3 mmol/L]**

Second measurement (two weeks later) : **[6.9 mmol/L]**

What is the probability likelihood that this is a real change?

- A: Unlikely (<50%)**
- B: Doubtful (50-80%)
- C: Likely (80-95%)
- D: Very likely(>95%)

### 2. TSH

**Reference values TSH [0.4-4.3 mU/L]**

First measurement: **[6.5 mU/L]**

Second measurement (three months later) : **[7.1 mU/L]**

What is the probability likelihood that this is a real change?

- A: Unlikely (<50%)**
- B: Doubtful (50-80%)
- C: Likely (80-95%)
- D: Very likely (>95%)

### 3. CRP

*Reference values CRP: <10 mg/L*

First measurement: [79 mg/L]  
Second measurement (two days later): [106 mg/L]

What is the probability likelihood that this is a real change?

- A: Unlikely (<50%)
- B: Doubtful (50-80%)
- C: Likely (80-95%)
- D: Very likely (>95%)**

### 4. Haemoglobin

*Reference values haemoglobin for a woman (7.5-9.5 mmol/L)*

First measurement: [7.0 mmol/L]  
Second measurement (six months later): [7.2 mmol/L]

What is the probability likelihood that this is a real change?

- A: Unlikely (<50%)**
- B: Doubtful (50-80%)
- C: Likely (80-95%)
- D: Very likely (>95%)

### 5. HbA1c

*Target value HbA1c <58 mmol/mol*

First measurement: [55 mmol/mol] (= 7.2%)  
Second measurement (6 months later): [60 mmol/mol] (= 7.6%)

What is the probability likelihood that this is a real change?

- A: Unlikely (<50%)
- B: Doubtful (50-80%)
- C: Likely (80-95%)**
- D: Very likely (>95%)

## 6. Leukocytes

*Reference values leukocytes [3.5-11.0\*10<sup>9</sup>/L]*

First measurement: [12\*10<sup>9</sup>/L]

Second measurement (14 days later): [14\*10<sup>9</sup>/L]

What is the probability likelihood that this is a real change?

A: Unlikely (<50%)

**B: Doubtful (50-80%)**

C: Likely (80-95%)

D: Very likely (>95%)

## 7. aPTT (this question was excluded from analyses)

*Reference values aPTT [18-36 sec]*

First measurement: [30 sec]

Second measurement (a few hours later) : [34 sec]

What is the probability likelihood that this is a real change?

A: Unlikely (<50%)

B: Doubtful (50-80%)

C: Likely (80-95%)

**D: Very likely (>95%)**

## 8. Vitamin B12

*Reference values vitamin B12 [250-850 pmol/L]*

First measurement: [114 pmol/L]

Second measurement (3 months later): [125 pmol/L]

What is the probability likelihood that this is a real change?

**A: Unlikely (<50%)**

B: Doubtful (50-80%)

C: Likely (80-95%)

D: Very likely (>95%)

## 9. ALAT

*Reference values ALAT < 45U/L*

First measurement: [44 U/L]

Second measurement (6 months later): [55 U/L]

What is the probability likelihood that this is a real change?

A: Unlikely (<50%)

**B: Doubtful (50-80%)**

C: Likely (80-95%)

D: Very likely (>95%)

## 10. Creatinine

*Reference values creatinine for men [60-110 µmol/L]*

First measurement: [89 µmol/L]

Second measurement (one year later): [109 µmol/L]

What is the probability likelihood that this is a real change?

A: Unlikely (<50%)

B: Doubtful (50-80%)

C: Likely (80-95%)

**D: Very likely (>95%)**

# BMJ Open

## ***Labtracker+*, a medical smartphone app for the interpretation of consecutive laboratory results: an external validation study**

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2017-015854.R1
Article Type:	Research
Date Submitted by the Author:	24-May-2017
Complete List of Authors:	Hilderink, Judith; Maastricht University Medical Center+, Central Diagnostic Laboratory, Clinical Chemistry Rennenberg, Roger; Maastricht University Medical Center+, Internal Medicine Vanmolkot, Floris; Maastricht University Medical Center+, Internal Medicine Bekers, Otto; Maastricht University Medical Center+, Central Diagnostic Laboratory, Clinical Chemistry Koopmans, Richard; Maastricht University Medical Center+, Internal Medicine Meex, Steven; Maastricht University Medical Center+, Central Diagnostic Laboratory, Clinical Chemistry
<b>Primary Subject Heading</b>:	Medical education and training
Secondary Subject Heading:	Diagnostics
Keywords:	Medical application, Biological variation, MEDICAL EDUCATION & TRAINING

SCHOLARONE™  
Manuscripts

**Full title**

*Labtracker+*, a medical smartphone app for the interpretation of consecutive laboratory results:  
an external validation study

**List of authors**

J.M. Hilderink<sup>1</sup>, [judith.hilderink@mumc.nl](mailto:judith.hilderink@mumc.nl)

R.J.M.W. Rennenberg<sup>2</sup>, [r.rennenberg@mumc.nl](mailto:r.rennenberg@mumc.nl)

F.H.M. Vanmolkot<sup>2</sup>, [floris.vanmolkot@mumc.nl](mailto:floris.vanmolkot@mumc.nl)

O. Bekers<sup>1</sup>, [o.bekers@mumc.nl](mailto:o.bekers@mumc.nl)

R.P. Koopmans<sup>2</sup>, [richard.koopmans@mumc.nl](mailto:richard.koopmans@mumc.nl)

S.J.R. Meex<sup>1</sup>, [steven.meex@mumc.nl](mailto:steven.meex@mumc.nl)

**Affiliations**

<sup>1</sup> Department of Clinical Chemistry, Central Diagnostic Laboratory, Maastricht University  
Medical Centre+, Maastricht, the Netherlands

<sup>2</sup> Department of Internal Medicine, Maastricht University Medical Centre+, Maastricht, the  
Netherlands

**Corresponding author**

Steven J. R. Meex

Department of Clinical Chemistry, Cardiovascular Research Institute Maastricht (CARIM),  
Maastricht University Medical Centre+ (MUMC+)

P.O. Box 5800, 6202 AZ Maastricht, the Netherlands

1  
2  
3 Telephone number: 31-(0)43-3874694  
4

5 Fax number: 31-(0)43-3874692  
6

7  
8 [steven.meex@mumc.nl](mailto:steven.meex@mumc.nl)  
9

10  
11  
12  
13 **Word count main text**

14  
15 2805  
16  
17

18  
19  
20 **Keywords**

21 Medical application, biological variation, medical education  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



## ABSTRACT

**Objectives:** When monitoring patients over time, clinicians may struggle to distinguish ‘real changes’ in consecutive blood parameters from so-called natural fluctuations. In practice, they have to do so by relying on their clinical experience and intuition. We developed *Labtracker+*, a medical app that calculates the probability that an increase or decrease over time in a specific blood parameter is real, given the time between measurements.

**Design:** We presented patient cases to 135 participants to examine whether there is a difference between medical students, residents and experienced clinicians when it comes to interpreting changes between consecutive laboratory results. Participants were asked to interpret if changes in consecutive laboratory values were likely to be ‘real or rather due to natural fluctuations. The answers of the study participants were compared to the calculated probabilities by the app *Labtracker+* and the concordance rates were assessed.

**Setting and participants:** Medical students ( $n=92$ ), medical residents from the department of internal medicine ( $n=19$ ) and internists ( $n=24$ ) at a Dutch university medical centre.

**Primary and secondary outcome measures:** Concordance rates between the study participants and the calculated probabilities by the app *Labtracker+* were compared. Besides, we tested whether physicians with clinical experience scored better concordance rates with the app *Labtracker+* than inexperienced clinicians.

**Results:** Medical residents and internists showed significantly better concordance rates with the calculated probabilities by the app *Labtracker+* than medical students, regarding their interpretation of differences between consecutive laboratory results ( $p=0.009$  and  $p<0.001$ , respectively).

1  
2  
3 **Conclusion:** The app *Labtracker+* could serve as a clinical decision tool in the interpretation of  
4 consecutive laboratory test results, and could contribute to rapid recognition of parameter  
5 changes by physicians.  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

## STRENGTHS AND LIMITATIONS OF THIS STUDY

- This study describes an objective scientific approach of distinguishing a ‘true change’ from a ‘random fluctuation’, when interpreting a change between two consecutive laboratory measurements
- Using *Labtracker+* could be useful in reducing information overload and enabling rapid recognition of parameter changes by expert clinicians
- The database that was used to develop the app *Labtracker+* presently contains only a limited number of laboratory parameters
- The cases we used to compare the concordance rates between *Labtracker+* calculations and clinicians’ estimates, were not presented in random order

## INTRODUCTION

Medical doctors frequently monitor the laboratory results of patients to determine whether their condition is stable, improving or deteriorating, and to decide whether treatment should be continued, initiated, or rather be postponed. Hence, an intuitive ‘reference frame’ of the patient can be formed by repeated assessment of laboratory parameters. An expertise-based reference frame is often lacking in less experienced clinicians, such as medical residents, or medical students. Hence, it can be difficult for less experienced clinicians to distinguish natural fluctuations in laboratory results from ‘real changes’ and vice versa.

According to a study of Jones et al., 1 in 5 medical graduates identified themselves as being ‘less than competent’ in using laboratory testing.<sup>1</sup> Additionally, a study of Hickner et al. demonstrated that even experienced clinicians reported uncertainty in interpreting laboratory test results in approximately 10% of the ordered laboratory tests.<sup>2</sup>

To help (inexperienced) clinicians to make the distinction whether a change between measurements is real or not, we developed a medical app (*Labtracker+*), that calculates the probability of a ‘real change’ between two consecutive laboratory results. It is simple to use in clinical practice and provides evidence-based decision support to complement intuitive interpretation of changes in consecutive laboratory results.

The current study tests the hypothesis that experienced clinicians show better concordance rates with *Labtracker+*-based calculations than medical students and less experienced clinicians.

## METHODS

This study was approved by the Dutch Association for Medical Education (NVMO). Study participants provided written informed consent.

### Study participants

135 participants were included: 92 medical students, 19 medical residents, and 24 medical specialists. We scheduled six days on which medical students participated in our study. All students were at the same stage of their internship and 92% of them were in their first medical internship (usually in their fourth year of the six years training to become a medical doctor in the Netherlands). The medical residents and medical specialists, all employed at the department of internal medicine at Maastricht University Medical Centre, were recruited during the daily morning briefing. The mean number of clinical experience years was  $24 \pm 9$  for the internist and  $3.5 \pm 2$  for the residents.

Exclusion criteria were: medical students in the bachelor phase of their studies and residents or specialists of a specialty other than internal medicine.

### Study design

Ten fictive clinical cases were presented to the study participants. These cases were conceived and developed by the authors of this article. Central to each case was to interpret the likelihood of change between consecutive measurements of a certain laboratory parameter. The time interval between the measurements varied for all cases. The cases can be found in the supplemental material.

1  
2  
3 The participants were asked to categorize the presented changes in one of four likelihoods:  
4  
5 ‘unlikely’ (<50%), ‘doubtful’ (50-80%), ‘probably’ (80-95%), and ‘very likely’ (> 95%).  
6  
7

8 Participants answered these questions individually. The extent to which the responses of the  
9  
10 participants were in accordance with the calculated probabilities by the app *Labtracker+* were  
11  
12 analysed.  
13

14  
15 The laboratory parameters that were incorporated in the cases were: haemoglobin, leukocytes,  
16  
17 thyroid stimulating hormone (TSH), vitamin B12, alanine aminotransferase (ALAT), glucose,  
18  
19 glycosylated haemoglobin type A1c (HbA1c), creatinine, activated partial thromboplastin time  
20  
21 (aPTT) and C-reactive protein (CRP).  
22  
23

### 24 25 26 27 28 **Statistical analysis**

29  
30 All answers given by the study participants were compared to the calculated probabilities by the  
31  
32 app *Labtracker+*. Two statistical analyses were performed. Firstly, differences among the study  
33  
34 groups were calculated using one-way analysis of variance (ANOVA). *p*-values < 0.05 were  
35  
36 considered statistically significant. Comparisons between pairs of groups were adjusted for  
37  
38 multiple testing using the Bonferroni correction.  
39

40  
41 Secondly, two groups were compared; the medical students formed the ‘inexperienced group’  
42  
43 and the medical residents and internists together formed the ‘experienced group’. The two groups  
44  
45 were compared using the Fisher’s exact test. All statistical calculations were performed using  
46  
47 SPSS version 23 (IBM SPSS Statistics. IBM Corporation, Armonk, New York).  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

### Calculating the probability of change using the app *Labtracker+*

For a change between consecutive measurements to become significant, the difference must be larger than the change that would reasonably be expected due to normal analytical- and biological variation. This is termed the reference change value (RCV) and was first described by Fraser et al.<sup>3</sup> The RCV has become an established concept in laboratory medicine.

The RCV differs for each parameter and depends on the within-subject biological variation ( $CV_I$ ) and the analytical variation ( $CV_A$ ) and can be calculated as follows:<sup>4-5</sup>

$$RCV = \sqrt{2} * Zscore * \sqrt{CV_A^2 + CV_I^2}$$

In this formula, the *Z*-score represents the number of standard deviations and corresponds to the desired probability. Commonly used *Z*-scores are 1.96 and 2.56. These *Z*-scores calculate the percentage increase or decrease that is required to become statistically significant, with a false positive rate of 5%, ( $p < 0.05$ ) and 1% ( $p < 0.01$ ), respectively.

An important side note to calculate the RCV is that not all clinical decisions are made with statistical probabilities of >95%. In other words, some changes are strictly not statistically significant, but the probability may be high enough to lead to a treatment or clinical decision. For example, a haemoglobin concentration of 6.1 mmol/L was measured in a male patient. After a week, it dropped to 5.8 mmol/L. Although the likelihood probability of a 'real' decrease in haemoglobin level is less than 95% in this case - namely around 70% - the clinician may still decide to use additional diagnostic testing or to treat the patient.

For implementation in the smartphone app, we rearranged the RCV-formula so that the *Z*-score (and therefore the likelihood probability) is the unknown. The probability of a real change between consecutive measurements will then be calculated. The rearranged RCV-formula is:

$$Zscore = \frac{\text{percentage change between consecutive measurements}}{\sqrt{2 * (CV_A^2 + CV_I^2)}}$$

The calculations of the app *Labtracker+* are based on the above formula. This makes it possible to compare two consecutive laboratory results and to calculate the probability of a real change.

### **Time-dependency of within-subject biological variation in some parameters**

In a number of laboratory parameters, the within-subject biological variation ( $CV_I$ ) changes over time. We systematically reviewed published biological variation studies of 106 laboratory tests and found a time-dependent increase of the biological variation for 21 parameters. For the laboratory parameters with sufficient information about their biological variation in the scientific literature, the associations between biological variation data and the time interval to which they apply, were modeled statistically and integrated into *Labtracker+*. This enabled the calculation of probabilities of a change between measurements, while accounting for the time-span between consecutive measurements.

### **Practical functionality of the app *Labtracker+***

In total, 106 parameters are available in *Labtracker+*. When using the app, users can choose the desired parameter out of a list that is displayed in alphabetical order. Subsequently, they have to choose the 'body liquid' in which the measurements were performed and, if relevant, whether the blood sampling regards a male or a female person. However, this question is asked only to display an indication of the reference values, but it does not affect the final calculation of the probability. Then the user fills in the previous and the recent value, as well as the time interval in



1  
2  
3 days between the measurements. Also the ‘a priori expectation’ of the change between the  
4  
5 measurements have to be chosen (‘rise’ for an expected increase, ‘fall’ for an expected decrease  
6  
7 or ‘either’ when there is no prior expectation). *Labtracker+* will then calculate the probability  
8  
9 that the change between the two measured values is a ‘real change’. In figure 1, screenshots of  
10  
11 the app were presented to visualise the functionality of the app.  
12  
13

14  
15 For a more detailed description of the practical use of *Labtracker+*, see the user manual on the  
16  
17 website (<http://www.labtracker.info/resources/Manual-Labtracker.pdf>).  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## RESULTS

Data collection was incomplete for two subjects (1.5%); it concerns two medical students who did not fill in all questions and therefore, they were excluded from all analyses.

### Psychometric analysis

We performed the 'Item Test Correlation ( $R_{IT}$ )' test, a psychometric analysis to check the distinctiveness of the case questions.<sup>6 7</sup> Briefly, the  $R_{IT}$  is a correlation coefficient which can vary between -1 and 1; a question with a  $R_{IT}$  score of 0 does not distinguish between good and poor performers. If the  $R_{IT}$  score is negative then the question may not be valid and should be removed.<sup>6 7</sup> The  $R_{IT}$  score of the aPTT-question had a negative value and therefore, this question was excluded from further analyses. The  $R_{IT}$  score of the other questions were all considered sufficient.

### Comparison of the app *Labtracker+* with the estimates of physicians

The correct answer was defined as the answer that matched exactly with the calculated probability by the app *Labtracker+*, i.e. the answer category containing the calculated probability percentage.

In figure 2, the estimates of the study groups were shown. Participants with extensive clinical experience showed higher concordance rates with *Labtracker+* calculations; with internists and the medical residents scoring significantly better than the medical students ( $p < 0.001$  and  $p = 0.021$ , respectively). The percentage of questions answered correctly by the medical residents did not differ significantly from the internists ( $p = 0.719$ ; 95% CI -5.4 – 15.5).

A sensitivity analysis was conducted that allowed a broader range to define concordant answers.

1  
2  
3 In this sensitivity analysis, not only the answer category that matched exactly with the calculated  
4 probability by *Labtracker+* was considered correct, but also the answer category adjacent to that  
5  
6 probability by *Labtracker+* was considered correct, but also the answer category adjacent to that  
7  
8 category. The results of this analysis are shown in figure 3. On average, the medical students  
9  
10 answered 69% of the questions correctly, the medical residents 78% and the internists 81%.  
11  
12 Similar to the primary analysis, concordance rates of answers from internists with the  
13  
14 *Labtracker*-calculated probabilities were significantly better than medical students ( $p<0.001$ ;  
15  
16 95% CI 4.7 – 19.0). Again, the difference between medical residents and medical students was  
17  
18 statistically significant as well ( $p=0.020$ ; 95% CI 1.1 – 16.8). And, like the first analysis,  
19  
20 concordance rates between the medical residents and *Labtracker+* did not differ significantly  
21  
22 from the internists ( $p=1.000$ ; 95% CI -12.5 – 6.6). Therefore, in subsequent analyses, internists  
23  
24 and medical residents were pooled as one group (the experienced clinicians) and compared with  
25  
26 medical students.  
27  
28  
29  
30  
31  
32  
33  
34

### 35 Analysis per parameter

36  
37 The scores of the study participants for each parameter were assessed separately. Figure 4  
38  
39 depicts the percentage of correct answers from the study groups per parameter. The  
40  
41 ‘experienced’ group (the residents and the internists) was compared with the medical students. In  
42  
43 general, the experienced group demonstrated the most concordant answers, compared to the  
44  
45 inexperienced medical students. The experienced clinicians scored best on the parameter  
46  
47 haemoglobin (95%) and the parameter they scored the worst on was creatinine (28%). The  
48  
49 medical students scored best on the haemoglobin parameter as well (82%). They scored worst on  
50  
51 the TSH question (6%).  
52  
53  
54

55 In accordance with the previous analysis, we conducted a sensitivity analysis in which the  
56  
57  
58  
59  
60

1  
2  
3 category adjacent to the correct answer was also considered to be correct. These results are  
4 shown in figure 5. Again, the experienced clinician gave the most concordant answers. In this  
5  
6 analysis, the experienced group (residents and internists) scored 88% correct answers on average.  
7  
8 They scored the best on the vitamin B12, leukocytes, and haemoglobin questions (98%). The  
9  
10 TSH question was the least well answered of all parameters (70%). The medical students scored  
11  
12 90% correct answers on the haemoglobin question. The least concordant question by the students  
13  
14 was TSH (44%).  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## DISCUSSION

The current study demonstrated that experienced clinicians show better concordance rates with the calculated probabilities by the app *Labtracker+* than less experienced medical students.

Stratified analyses reveal that this pattern holds true for the majority of studied laboratory parameters. Concordance rates with the app demonstrate an incremental improvement from medical students, to residents and internists. The differences between clinicians with experience (internists and residents) and inexperienced medical students were statistically significant. This outcome illustrates the relevance and added value of the medical app *Labtracker+*.

The clear association between the degree of clinical experience and improved concordance rates with the calculated probabilities by the app supports the validity of *Labtracker+* calculations.

Improved concordance rates with increasing clinical experience suggest that *Labtracker+* may well be of benefit for less experienced clinicians, such as medical students.

However, *Labtracker+* is not only a useful tool for inexperienced clinicians. We believe the app can form a 'bridge' between laboratory science and clinical practice. *Labtracker+* could be useful in reducing information overload and enabling rapid recognition of parameter changes by expert clinicians, especially in time constrained and rapidly changing environments (e.g., intensive care units). Clinicians do not have to make cumbersome calculations when estimating whether a change between consecutive measurements is a real change or not. By using the app, the quality of the laboratory information that often plays an important role in clinical decisions will improve. *Labtracker+* enables clinicians to make efficient and standardized interpretations with minimal cognitive effort. Of course, the application will never replace clinical judgement, but it can serve as a supportive aid for diagnosis and treatment.

1  
2  
3 In addition, in current clinical practice, some clinical tasks that were always performed by  
4 physicians now more and more become the responsibility of physician assistants or nurse  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

In addition, in current clinical practice, some clinical tasks that were always performed by physicians now more and more become the responsibility of physician assistants or nurse specialists. Such ‘shifts of tasks’ can be for instance, a diabetic nurse who monitors the routine diabetic check-ups. In 2014, Lenters - Westra et al. demonstrated that there are significant differences in interpretation of changes in HbA1c results between physicians and nurses <sup>8</sup>.

*Labtracker+* may serve as a useful support to standardize care in such a situation.

The benefit offered by the app may be parameter dependent. According to this external validation study, the concordance rates with the calculated probabilities varied from 6% for students for the TSH question to 98% for experienced clinicians for the haemoglobin question. This difference in concordance rates suggests that the ‘number intuition’ of a clinician differs per laboratory parameter and varies with clinical experience. It may be difficult for a clinician to estimate how a ‘normal value’ of a parameter fluctuates over time in an individual if the clinician has little actual experience with that parameter. In other words, a clinician with more experience in interpreting a specific parameter can better estimate its biological variation. Since the scores of the medical residents were not significantly different from those of the internists, it might be assumed that medical residents develop their ‘number intuition’ early in the residency training. The participants scored not all parameters equally concordant. A potential explanation for the ‘worse scored questions’ could be that not all laboratory parameters are widely used in clinical practice, and that clinicians therefore have little experience in interpreting those parameters.

Another potential explanation could be that not all questions were of equal difficulty.

Some limitations in the current study merit attention. First, the groups of medical residents and specialists were relatively small ( $n=19$  and  $n=24$ , respectively). Second, only nine parameters

1  
2  
3 were tested, with one single interpretation per parameter. However, our estimation is that the  
4  
5 study groups are a good reflection of the total population of people in these groups. In addition,  
6  
7 the parameters we studied are commonly used in clinical practice and cover a broad range of  
8  
9 clinical domains. Therefore, it can reasonably be expected that the results of this study can be  
10  
11 generalized to all commonly used laboratory parameters. Third, the cases were not presented in  
12  
13 random order to the subjects. It is possible that the subjects' performances were partially  
14  
15 influenced by the order of case presentation. However, if there were a learning curve, then this  
16  
17 should be the same for all three study groups.  
18  
19

20  
21 Fourth, it remains unclear whether the use of the medical app *Labtracker+*, would affect clinical  
22  
23 decision-making. In that respect, a study that incorporated medical decision-making, based on  
24  
25 hypothetical cases involving laboratory test result interpretation(s), would be an interesting  
26  
27 addition. The hypothesis that could then be tested is whether inexperienced clinicians make the  
28  
29 same decisions as more experienced clinicians when they use the app *Labtracker+*.  
30  
31  
32  
33  
34  
35

36  
37 In 2014, Abbot et al. demonstrated that the self-assessed knowledge about interpreting laboratory  
38  
39 results of inexperienced medical residents significantly improved after receiving an educational  
40  
41 program about 'the basics of laboratory test ordering and interpretation'<sup>9</sup>. This finding endorses  
42  
43 the importance of education for inexperienced clinicians. Elaborating on this, it would be  
44  
45 interesting to investigate whether *Labtracker+*, as an educative aid, could provide better medical  
46  
47 decision-making.  
48  
49

50  
51 In conclusion, the current study has provided insight into the intuition of medical students with  
52  
53 hardly any clinical experience, clinicians with some experience (medical residents from the  
54  
55 department of internal medicine), and experienced clinicians (internists), when interpreting a  
56  
57  
58  
59  
60

1  
2  
3 difference between consecutive laboratory results The findings suggest the existence of a trend  
4  
5 that the ‘number intuition’ of a clinician improves with more years of experience. The  
6  
7 development of this intuition fosters a better interpretation of laboratory results and could  
8  
9 possible improve clinical decision-making. In that sense, the use of the medical app *Labtracker+*  
10  
11 can be a useful tool, especially for inexperienced clinicians. Besides, *Labtracker+* could be  
12  
13 useful in reducing information overload and enabling rapid recognition of parameter changes by  
14  
15 expert clinicians.  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



## REFERENCES

1. Jones A, McArdle PJ, O'Neill PA. Perceptions of how well graduates are prepared for the role of pre-registration house officer: a comparison of outcomes from a traditional and an integrated PBL curriculum. *Med Educ* 2002;**36**(1):16-25.
2. Hickner J, Thompson PJ, Wilkinson T, et al. Primary care physicians' challenges in ordering clinical laboratory tests and interpreting results. *J Am Board Fam Med* 2014;**27**(2):268-74.
3. Fraser CG. Reference change values. *Clin Chem Lab Med* 2012;**50**(5):807-12.
4. Ricos C, Alvarez V, Perich C, et al. Rationale for using data on biological variation. *Clin Chem Lab Med* 2015;**53**(6):863-70.
5. Ricos C, Alvarez V, Cava F, et al. Current databases on biological variation: pros, cons and progress. *Scand J Clin Lab Invest* 1999;**59**(7):491-500.
6. Ebel RL, Frisbie DA. *Essentials of Educational Measurement*. Pearson, 5th edition 1991.
7. Ory JC, Ryan KE. *Tips for Improving Testing and Grading*. Sage Publications Inc 1993.
8. Lenters-Westra E, Schindhelm RK, Bilo HJ, et al. Differences in interpretation of haemoglobin A1c values among diabetes care professionals. *Neth J Med* 2014;**72**(9):462-6.
9. Abbott M, Paulin H, Sidhu D, et al. Laboratory tests, interpretation, and use of resources: a program to introduce the basics. *Can Fam Physician* 2014;**60**(3):e167-72.

## AUTHOR'S CONTRIBUTION

Concept and design of the study, interpretation of the data and drafting of the manuscript: JMH, SJRM and RPK. Statistical analysis: JMH. Critically revising the manuscript and providing important intellectual content: RJMWR, FHMV, OB.

## FUNDING

This work was supported by an Academic Incentive Research Grant from Maastricht University Medical Center+ to dr. Meex.

## COMPETING INTERESTS

*Labtracker+* was conceived and developed by J. Hilderink, R. Koopmans, R. Rennenberg, and S. Meex. It is a free medical app, and therefore financial interests do not apply.

## DATA SHARING STATEMENT

There is no additional unpublished data available.

## FIGURE CAPTIONS

### Figure 1:

Title: Screenshots of the app *Labtracker+*

### Figure 2

Title: Percentage concordant answers per group

Caption: Values are mean percentages  $\pm$  SD.

Medical students vs. internists,  $p < 0.001$ ; 95% CI 6.96–22.59, medical students vs. medical residents,  $p = 0.009$ ; 95% CI 1.10–18.28, medical residents vs. internists,  $p = 0.321$ ; 95% CI -5.36–15.53.

$p$ -values  $< 0.05$  were considered statistically significant (\*).

### Figure 3:

Title: Percentage concordant answers when also the probability adjacent to the designated category was considered correct.

Caption: Values are mean percentages  $\pm$  SD.

Medical students vs. internists,  $p = 0.007$ ; 95% CI 4.72–19.00, medical students vs. medical residents,  $p = 0.031$ ; 95% CI 1.07–16.76, medical residents vs. internists,  $p = 0.609$ ; 95% CI -5.36–15.53.

$p$ -values  $< 0.05$  were considered statistically significant (\*).

**Figure 4:**

Title: Percentage concordant answers per parameter

Caption: TSH  $p < 0.001$ , creatinine  $p = 0.012$ , vitamin B12  $p = 0.024$ , HbA1c  $p = 0.036$ , haemoglobin  $p = 0.056$ , ALAT  $p = 0.237$ , leukocytes  $p = 0.307$ , glucose  $p = 0.462$ , CRP  $p = 0.574$

In this figure, the medical residents and the internists were combined into the group 'clinicians with experience'

**Figure 5:**

Title: Percentage concordant answers per parameter when also the probability adjacent to the designated category was considered correct.

Caption: Vitamin B12  $p = 0.002$ , TSH  $p = 0.006$ , leukocytes  $p = 0.012$ , haemoglobin  $p = 0.167$ , creatinine  $p = 0.012$ , ALAT  $p = 0.221$ , glucose  $p = 0.329$ , HbA1c  $p = 1.000$ , CRP  $p = 1.000$ .

In this figure, the medical residents and the internists were combined into the group 'clinicians with experience'

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

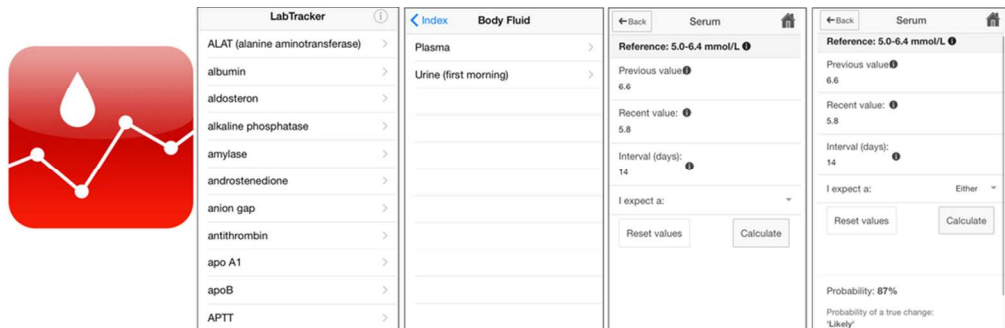


Figure 1: Screenshots of the app *Labtracker+*

88x28mm (300 x 300 DPI)

peer review only

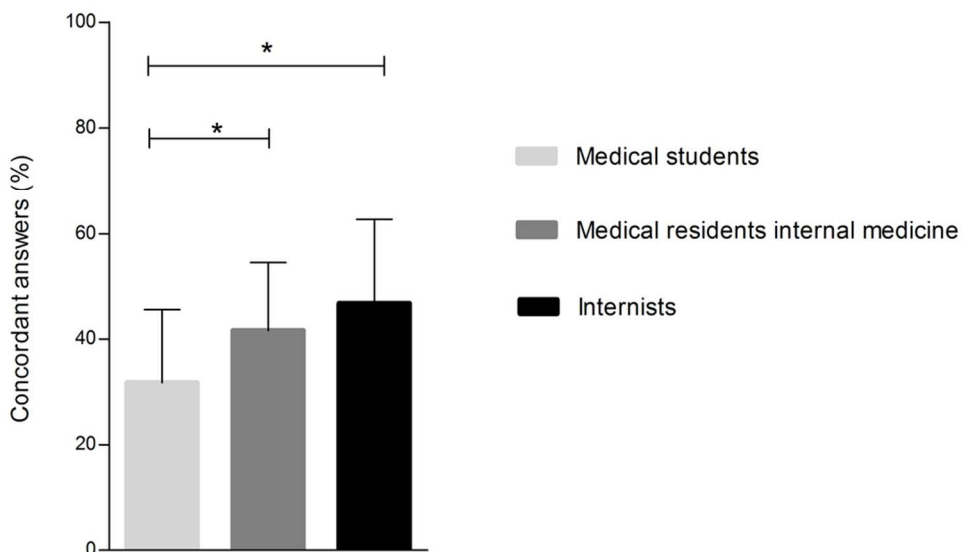


Figure 2:  
Title: Percentage concordant answers per group

Caption: Values are mean percentages  $\pm$  SD.  
 Medical students vs. internists,  $p < 0.001$ ; 95%CI 6.96–22.59, medical students vs. medical residents,  $p = 0.009$ ; 95% CI 1.10–18.28, medical residents vs. internists,  $p = 0.321$ ; 95% CI -5.36–15.53.  
 p-values  $< 0.05$  were considered statistically significant (\*).

92x57mm (300 x 300 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

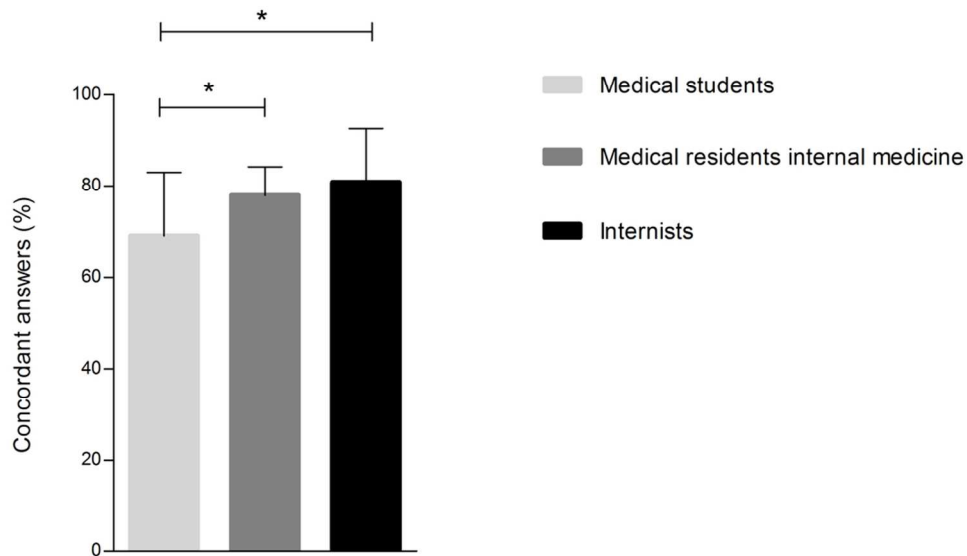


Figure 3:

Title: Percentage concordant answers when also the probability adjacent to the designated category was considered correct.

Caption: Values are mean percentages  $\pm$  SD.

Medical students vs. internists,  $p=0.007$ ; 95% CI 4.72–19.00, medical students vs. medical residents,  $p=0.031$ ; 95% CI 1.07–16.76, medical residents vs. internists,  $p=0.609$ ; 95% CI -5.36–15.53.

$p$ -values  $< 0.05$  were considered statistically significant (\*)

96x60mm (300 x 300 DPI)

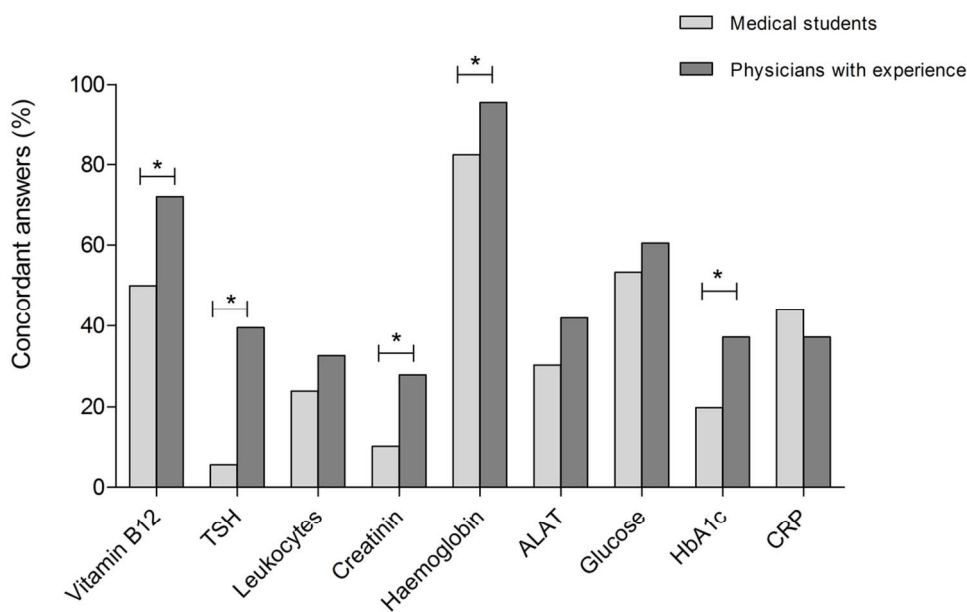


Figure 4:  
Title: Percentage concordant answers per parameter

Caption: TSH  $p < 0.001$ , creatinine  $p = 0.012$ , vitamin B12  $p = 0.024$ , HbA1c  $p = 0.036$ , haemoglobin  $p = 0.056$ , ALAT  $p = 0.237$ , leukocytes  $p = 0.307$ , glucose  $p = 0.462$ , CRP  $p = 0.574$   
In this figure, the medical residents and the internists were combined into the group 'clinicians with experience'

108x69mm (300 x 300 DPI)



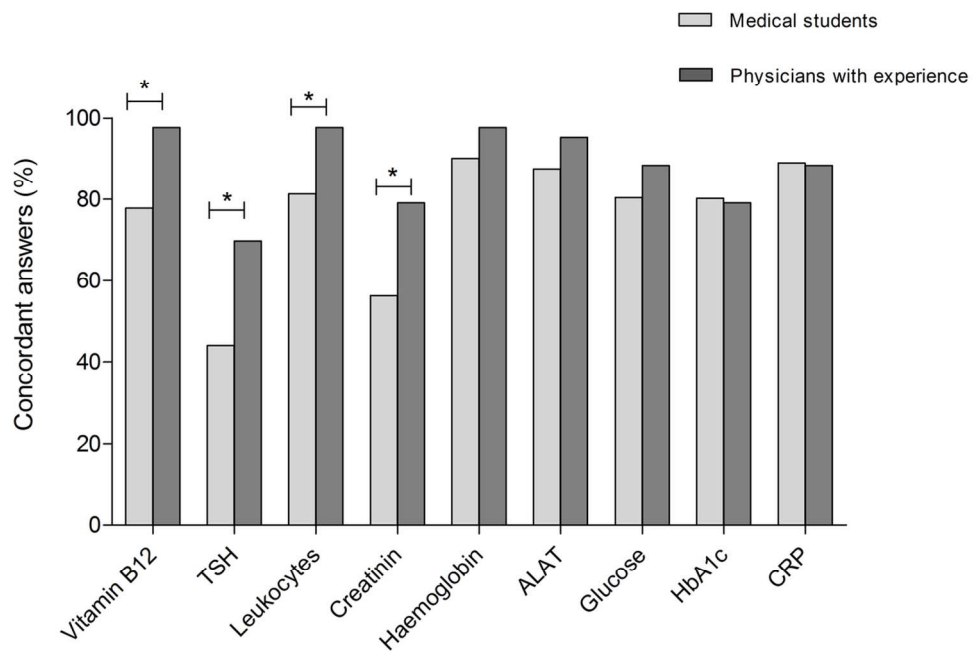


Figure 5:

Title: Percentage concordant answers per parameter when also the probability adjacent to the designated category was considered correct.

Caption: Vitamin B12  $p=0.002$ , TSH  $p=0.006$ , leukocytes  $p=0.012$ , haemoglobin  $p=0.167$ , creatinine  $p=0.012$ , ALAT  $p=0.221$ , glucose  $p=0.329$ , HbA1c  $p=1.000$ , CRP  $p=1.000$ .

In this figure, the medical residents and the internists were combined into the group 'clinicians with experience'

114x78mm (300 x 300 DPI)

## Supplementary material: Case questions

Correct answers, calculated by the app *Labtracker+*, are shown in bold.

Reference values are from Maastricht University Medical Center (MUMC+).

### 1. Glucose

**Reference values fasting glucose [4.0 – 7.0 mmol/L]**

First measurement: **[7.3 mmol/L]**

Second measurement (two weeks later) : **[6.9 mmol/L]**

What is the probability likelihood that this is a real change?

- A: Unlikely (<50%)**
- B: Doubtful (50-80%)
- C: Likely (80-95%)
- D: Very likely(>95%)

### 2. TSH

**Reference values TSH [0.4-4.3 mU/L]**

First measurement: **[6.5 mU/L]**

Second measurement (three months later) : **[7.1 mU/L]**

What is the probability likelihood that this is a real change?

- A: Unlikely (<50%)**
- B: Doubtful (50-80%)
- C: Likely (80-95%)
- D: Very likely (>95%)

### 3. CRP

*Reference values CRP: <10 mg/L*

First measurement: [79 mg/L]  
 Second measurement (two days later): [106 mg/L]

What is the probability likelihood that this is a real change?

- A: Unlikely (<50%)
- B: Doubtful (50-80%)
- C: Likely (80-95%)
- D: Very likely (>95%)**

### 4. Haemoglobin

*Reference values haemoglobin for a woman (7.5-9.5 mmol/L)*

First measurement: [7.0 mmol/L]  
 Second measurement (six months later): [7.2 mmol/L]

What is the probability likelihood that this is a real change?

- A: Unlikely (<50%)**
- B: Doubtful (50-80%)
- C: Likely (80-95%)
- D: Very likely (>95%)

### 5. HbA1c

*Target value HbA1c <58 mmol/mol*

First measurement: [55 mmol/mol] (= 7.2%)  
 Second measurement (6 months later): [60 mmol/mol] (= 7.6%)

What is the probability likelihood that this is a real change?

- A: Unlikely (<50%)
- B: Doubtful (50-80%)
- C: Likely (80-95%)**
- D: Very likely (>95%)

## 6. Leukocytes

*Reference values leukocytes [3.5-11.0\*10<sup>9</sup>/L]*

First measurement: [12\*10<sup>9</sup>/L]

Second measurement (14 days later): [14\*10<sup>9</sup>/L]

What is the probability likelihood that this is a real change?

A: Unlikely (<50%)

**B: Doubtful (50-80%)**

C: Likely (80-95%)

D: Very likely (>95%)

## 7. aPTT (this question was excluded from analyses)

*Reference values aPTT [18-36 sec]*

First measurement: [30 sec]

Second measurement (a few hours later) : [34 sec]

What is the probability likelihood that this is a real change?

A: Unlikely (<50%)

B: Doubtful (50-80%)

C: Likely (80-95%)

**D: Very likely (>95%)**

## 8. Vitamin B12

*Reference values vitamin B12 [250-850 pmol/L]*

First measurement: [114 pmol/L]

Second measurement (3 months later): [125 pmol/L]

What is the probability likelihood that this is a real change?

**A: Unlikely (<50%)**

B: Doubtful (50-80%)

C: Likely (80-95%)

D: Very likely (>95%)

## 9. ALAT

*Reference values ALAT < 45U/L*

First measurement: [44 U/L]

Second measurement (6 months later): [55 U/L]

What is the probability likelihood that this is a real change?

A: Unlikely (<50%)

**B: Doubtful (50-80%)**

C: Likely (80-95%)

D: Very likely (>95%)

## 10. Creatinine

*Reference values creatinine for men [60-110 µmol/L]*

First measurement: [89 µmol/L]

Second measurement (one year later): [109 µmol/L]

What is the probability likelihood that this is a real change?

A: Unlikely (<50%)

B: Doubtful (50-80%)

C: Likely (80-95%)

**D: Very likely (>95%)**