Hidden in Plain Sight

Overlooked Results and Other Errors in Evaluating Online Laboratory Results

Monkman, Helen; MacDonald, Leah; Nøhr, Christian Gradhandt; Tanaka, James W.; Lesselroth, Blake J.

Published in:
MEDINFO 2021

DOI (link to publication from Publisher):
10.3233/SHTI220203

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

MEDINFO 2021: One World, One Health – Global Partnership for Digital Innovation
P. Otero et al. (Eds.)

doi:10.3233/SHTI220203

Hidden in Plain Sight: Overlooked Results and Other Errors in Evaluating Online Laboratory Results

Helen Monkman[a], Leah MacDonald[a], Christian Nøhr[b], James W. Tanaka[c], and Blake J. Lesselroth[d]

[a] *School of Health Information Science, University of Victoria, Victoria, British Columbia, Canada*
[b] *Mærsk Mc-Kinney Møller Institute, University of Southern Denmark, Odense, Denmark*
[c] *Department of Psychology, University of Victoria, Victoria, British Columbia, Canada*
[d] *University of Oklahoma-Tulsa, School of Community Medicine, Tulsa, Oklahoma, USA*

Abstract

People are increasingly accessing their own laboratory (lab) results online. However, Canadians may be expected to use different systems to access their results, depending upon where they are tested (e.g., community lab vs. hospital), and these results may be displayed differently. This study examined the extent to which participants without medical expertise (N = 25) made errors identifying lab results (i.e., missing or mis-identifying abnormal results) in a mock report. Six participants overlooked each of the flagged values, 20 participants missed an abnormal result that was not flagged, and 2 participants mis-identified a normal value as out of range. We describe potential causes of these errors and the implications for the design of consumer-facing lab results.

Keywords:

Consumer Health Information, User-Centered Design, Clinical Laboratory Information Systems

Introduction

With the implementation of new consumer-facing health information technologies, people have greater access to their personal health information and medical records. One example of this is the ability to review laboratory (lab) results (e.g., COVID-19 tests, Pap smears, blood work) using online portals. Increased access to our own lab results may translate to improved decision making, self-management, and preparation during encounters with health care providers [1]. Many Canadians have access to their online lab results for lab tests collected in the community. However, lab tests collected in hospitals are increasingly available to Canadians, albeit using different online portals. Not only does this create an issue with information fragmentation, but it also means that people have to contend with different information systems, reporting formats, and data displays. In this study, we are interested in measuring the effect of different lab display configurations upon users' ability to detect abnormal or concerning values (i.e., beyond their reference range or described as notable).

Zikmund-Fisher and colleagues showed that health literacy and numeracy influenced the ability of patients to identify abnormal results presented in a table [1]. However, the reports used in their study did not include out-of-range flags; participants had to compare each value to its respective reference range to identify abnormal values [1]. Typically, online lab results give some indication (i.e., "flags") when values are outside of their respective reference ranges. However, how flags are displayed (e.g., colours, letters, bolding) often varies between systems. Further, tabular displays may not effectively communicate the magnitude of an abnormal finding or draw user attention to the most important issues [2]. By contrast, graphical depictions of abnormal results – including iconic representations of magnitude – can help participants recognize and prioritize relative urgency [2].

Although flags are undoubtedly beneficial for highlighting concerning results, some results may not include reference ranges or flags. Nonetheless, these results may still warrant attention. For example, tests with a binary outcome (i.e., positive or negative) or a qualitative description may not include a reference range and might not be flagged regardless of whether their outcome is typical or atypical.

In this study, we sought to determine the accuracy of users to identify abnormal values in a mock online lab report. In the report, we denoted flagged or out of range values slightly differently than the actual Canadian provincial portal. However, the display format was otherwise nearly identical.

Citizens in our sample had all used the provincial portal. Therefore, we hypothesized that they would have little difficulty accurately detecting flagged results. We also hypothesized that abnormal results missing a flag would be easily missed.

The process of finding information in a visual display can be summarized as a two step process:

> "Step 1. A visual query is formulated in the mind of the person, relating to the problem to be solved.
>
> Step 2. A visual search of the display is carried out to find patterns that resolve the query." (p.139) [3].

This study assisted with step 1 by asking participants to identify out of range results. As in the real world, not all abnormal results in the mock report were flagged (e.g., labs without a common reference range or labs with qualitative results).

Methods

The University of Victoria's Human Research Ethics Board approved this study. To recruit participants, we posted an invitation on a provincial (i.e., British Columbian) online platform for health research volunteers. Participants were eligible to participate if they were 19 years of age or older, had previously used online lab results, and neither practicing as, nor studying to be, a health care professional. Participants completed a questionnaire designed to collect demographic information and descriptions of experiences using online lab portals. One of the authors (HM) interviewed participants that consented to participate. Participants were offered $20 CAD for completing both the questionnaire and the interview.

Stimulus – Mock Online Lab Report Display

We only tested performance on a single stimulus within a mock online report with fictional data. The mock report was nearly identical (i.e., same columns and content) to the provincial community portal, with one important exception: flags denoting abnormal values (i.e., those below or above the reference range) were different. In the provincial portal, abnormal results are indicated with the letter "A" in the "flags" column. In our stimulus, high values were flagged with the letters "HI".

The stimulus contained 16 hematology, 8 urinalysis, 9 general chemistry, 6 urine chemistry, 2 complement testing, 1 immunology, and 2 serology test results. There were three elevated general chemistry results flagged with "HI": hemoglobin A1C/total hemoglobin (value = 6.0; reference range = <6.0%), potassium (value = 5.3; reference range = 3.5 – 5.2 mmol/L), and creatinine (value = 124; reference range = 67 – 117 umol/L). Also inlcuded was an estimated glomerular filtration rate (eGFR) of 53. While this value did not include a reference range or flag, language below the result read: "An eGFR result of 45-59 ml/min/1.73 m2 is consistent with mild to moderately decreased kidney function".

Procedure

We interviewed participants individually using Zoom videoconferencing software (Zoom Video Communications, Inc.). During the interviews, we furnished participants with the stimulus using the screen sharing feature and asked, "Are there any values outside the reference ranges?". The stimulus did not fit on a single screen; the lead investigtor [HM] scrolled or magnified the display at participants' requests. If participants were vague in their response (e.g., "that one is out of range") the interviewer asked them to clarify which test they meant. As in real world scenarios, participants could take as long as they wanted on this task.

Analysis

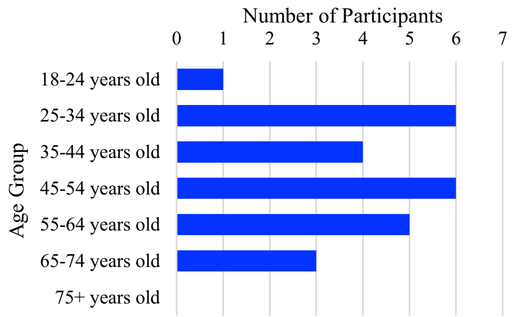
The research team analyzed participants' verbal responses to determine if a) they correctly identified the three abnormal results; b) they detected that the eGFR was abnormally low; and c) they identified any normal values as out of range (i.e., false postitives). We recorded data in a spreadsheet to analyze error rates. Each correct identification was coded with a 0; failure to identify (i.e. no mention) an abnormal result was coded with a 1. False positives were also coded with a 1. Additionally, we opportunistically recorded participant quotes related to the task. We summed the number of errors each participant made as well as for each type of error.

Results

Descriptive Statistics of the Sample

We included 25 participants in the analysis. There were twenty-three women and 20 born in Canada. All 25 participants primarily spoke English at home. Most participants were well educated; 23 had at least a post-secondary certificate or degre. As can be seen in **Figure 1**, with the exception of the highest and lowest age categories, participants were fairly evenly distributed in age.

Figure 1 – Distribution of Participants by Age Category



All 25 participant had been been accessing their online lab results for a minimum of 2 years. Nineteen of the 25 reported looking at their lab results "several times" a year, but five used them more often, and one less often. Nearly all participants (i.e., 24) reported using the provincial community lab portal, but five reported using other online lab results portals.

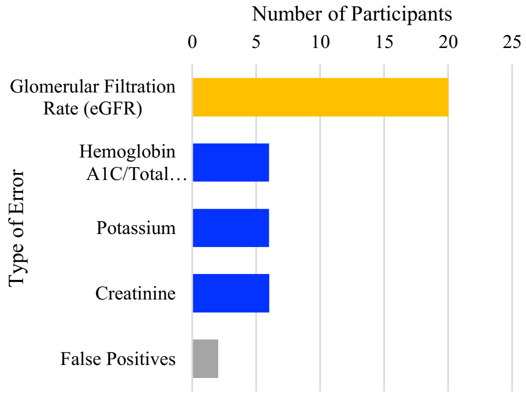
Error Rates for Abnormal Results

Only two participants did not make any errors reviewign results. That is, two people identified all three flagged results, the abnormal eGFR, and did not identify any false positive results. The remaining 23 participants made one or more errors.

The most frequent error was failing to recognize an abnormally low eGFR. Only five participants identified the eGFR result as abnormal; the remaining 20 did not (see **Figure 2**).

Participants were better at identifying flagged results, but we noted considerable variation in performance. Fifteen participants identified all three flagged results; the remainder failed to detect one or more flagged results. Three participants did not detect any flagged results and seven identified only one or two of the flagged results. Each abnormal, flagged result was missed by six particpants (see **Figure 2**). Two participants identified a normal result without any flag as out of range.

Figure 2 – Errors by Error Type: Number of Participants Who Missed Each Abnormal Result or Identified a False Positive



Display Weaknesses

Multiple participants described weaknesses with the display of the online lab results that may have affected perception and accuracy.

Table 1 – Quotes Illustrating Display Weaknesses

Problem	Quote
Column headers not visible in reports longer than the screen	Participant 20: "If I didn't remember at the top, what the different columns we're now seeing that it looks like there are three things that are high. So I'm assuming those are still part of that flag column, but it might be useful to have more headers, I guess."
Magnititude of abnormal value and urgency	Participant 11: "5.3 is a tad up – potassium. Is it serious? Probably not. But I should probably pay attention so should my doctor."
	Participant 21: "No, is that high? Does that mean that I have high hemoglobin? pre diabetes? Do I have diabetes? I don't know. Now I'm like nervous that I have diabetes."
Amount of information	Participant 17: "Yeah, see, there's a lot of information there that I don't think the average consumer needs. I think it's a lot of medical terminology that is just confusing to people."
Flagged results lack emphasis, priority, or clinical severity	Participant 15: "A color coordination system, this would be much easier. You could just like, go through your like, all of these are, this is green light, I'll go all the way down.
Lack of reference range	Participant 5: "The eGFR has no reference range. So we don't know if it's low or high."
Abnormal results not flagged	Participant 17: "And so they've got the note there, an eGFR result, blah, blah, blah, is consistent with mild to moderately decreased kidney function. So then I have to go back and find the eGFR, which is just above. Okay, so it's 53. So, I'm okay. But when I read that, I would think, wait a minute, do I have mild to moderately decreased kidney function? Why are they telling me that? That's good information, if it was within that? Oh, actually, no, I am in the range. So I do have mild – see, I was already confused. Yeah, so it's not clear."
Ambiguity of messaging	Participant 21: But also it wasn't flagged so well. I guess it was it says hi. [It would] be better if it was highlighted or like had like a little color or something."

Discussion

Many participants failed to identify abnormal results in the stimulus – even when given unlimited time to complete the task. Performance was especially poor interpreting eGFR values. Considering that flagged results are often the most important results for both patients and providers, they should be easy to identify. Unfortunately, our study suggests that abnormal results often go unnoticed.

We identified several reasons in our qualitative data why particpants may have failed to identify abnormal values. Additionally, we can speculate on other root causes using what we know about universal principles of design .

1. **Visibility:** Column headers for the results tables are static and only appear at the top of the report. Therefore, patients may not be able to read the headers when scrolling through longer reports.
2. **Cognitive load:** People may have more challenges recognizing abnormal results in longer reports with more descriptive text. The mock report was long and may have created a high cognitive load.
3. **Highlighting:** Flagged results are not clearly highlighted or prioritized by magnitude or urgency. Using letters rather than icons or colour coding to denote abnormal values may require more mental effort to detect.
4. **Figure-to-ground:** The flags are comprised of text and can be less obvious when adjacent to descriptive text. Essentially, the information blends into the background.
5. **Priming:** Although the stimulus was formatted to look very similar to the portal participants actually use, in a real world scenario, users would be alerted to the presence of at least one abnormal value in the report before opening the report itself. We did not include this in our study.
6. **Familiarity:** Participants may have been expecting the "A" flag they are familiar with because the report was nearly identical to the provincial lab portal they usually used.
7. **Magnitude:** Two of the test values (i.e., Hemoglobin A1C/Total Hemoglobin and Potassium) in this stimulus may have been challenging to detect because they fell just outside of the reference range. As observed in **Table 1**, participants had difficulty interpreting the urgency of results given their proximity to the reference range. The magnitude of the difference compared to a reference standard can be difficult to communicate in a tabular format. Graphical formats have been shown to communicate magnitude and its associated urgency more effectively [2].

Failing to Flag Important Results

The eGFR description provided a range for mild to moderately decreased kidney function. However, the abnormmal eGFR value was formatted differently from most other results; there was no flag, adjaceent reference range, or units. Patients indicated this caused confusion. Typically, reference ranges communicate the normal range. Therefore, it can be difficult for non-clinical users to recognize when the ranges reflect degrees of impairment or classes of abnormals (e.g., creatinine clearance for renal insufficiency or ejection fraction for heart failure) [4]. Perhaps a better way to the eGFR information would have been to include the normal range eGFR (i.e., > 90) and then the ranges for impairment. This approach is already used to report hemoglobin A1C/total hemoglobin values.

Inattentional Blindness

The visual communication and display issues outlined above may have collectively contributed to innattentional blindness. Inattentional blindness is "the failure to see highly visible objects we may be looking at directly when our attention is elsewhere" (p. 180) [5]. Inattentional blindness has been observed in a variety of situations from flight simulations where pilots failed to notice obstructions on the runway [6] to the famous "gorillas in our midst" experiment [7].

There are some important differences, however, between our study and previous inattentional blindness studies. Usually, the things that go unnoticed are unexpected. However, abnormal lab values are relatively common and should be expected. Also, inattentional blindness studies are typically dual or multi-task experiments. By contrast, in this study, errors of perception related to the *primary* and *sole* task we assigned to participants.

Preattentive Attributes

Preattentive attributes are cues or signals processed so quickly by the brain that they do not require conscious effort to be noticed [8]. In a printed document or electronic display, they draw our attention towards specific – and often critical – information . Preattentive processing occurs when preattentive attributes of the target stimulus differentiate it from distractors [9,10]. There are four main categories of preattentive visual attributes (features, or properties) that make stimuli almost immediately noticeable: colour, form, movement, and spatial positioning [8].

Without the aid of preattentive attributes, people must expend additional mental effort to process items serially (i.e., one after another) [9]. The presence of any distractors quickly compounds the complexity of the task; studies suggest detection time increases linearly with the number of distractors [3]. That is, the more distractors there are, the longer it takes for people to detect the target stimulus in a serial visual search. However, when preattentive attributes are present, people can use parallel procesing to more efficiently detect the target, regardless of the number of distractors [3].

Most real-world consumer-facing lab reports force users to adopt a serial focused attention approach when looking for flagged results. The inclusion of values that fall within reference standards function as distractors. Therefore, the number of distractors varies from zero to many depending upon how many values are outside (targets) or within (distractors) their respective reference ranges.

We believe purposeful application of preattentive attributes when designing lab reports could significantly improve patient attention, recognition, and understanding of medical information. For example, using graphical preattentive attributes for abnormal values would make them easy to see "at a glance"; thereby reducing the patient's cognitive load.

Recommendations for Online Lab Results

Ware (p. 14) [10] proposed the following guideline for the design of information visualizations:

> "Important data should be represented by graphical elements that are more visually distinct than those representing less important information."

This especially applies to design of laboratory reports. Abnormal results should be visually distinct and easy to recognize. Rather than only flagging results falling outside a reference range, all clinically abnormal results (e.g., results without a reference range) should be emphasized using common terms and visual iconography. Researchers have published recommendations [11] to improve the readability of online lab results. We believe these recommendations should be consistently used in system designs.

Limitations

There were several limitations to this study. First, we did not compare performance between the mock report and the actual display in the community portal. Therefore, we do not know how study findings compare to actual standards and performance. However, we suspect participants fared worse in our study because they were unfamiliar with our flagging notation. This fact notwithstanding, many of the design problems in this stimulus are also present in the standard display (e.g., hidden column headers; buried flags).

Second, our interview strategy may not have clearly directed participants to the eGFR issue. We asked participants to identify "out of range" results rather than ones that were "abnormal". Arguably, this simulates real life: people without clinical expertise are primarly concerned with identifying out of range values and may overlook other relevant data. Further, we did not ask people to perform this task as quickly and accurately as possible nor did we time their responses.

Third, this study included a small sample of well-educated, predominantly female users experienced reading online lab results. Given the characteristics of our sample, we expect that our participants would likely outperform those with less experience, education, or other factors. For example, we had no elderly participants who often face more challenges (1) due to the increased likelihood of comorbidities [12]; (2) using technology, given their numeracy and technology skills [13] as well as small interface elements (e.g., buttons, font size); and (3) understanding health information [14]. This issue deserves additional research.

Conclusions

People appreciate access to their results online [15] and Canadians are no exception [16]. As previously described, these portals are lauded for their potential to support patient decision-making, improve patient self-efficacy, and encourage richer conversations with providers [1]. However, this study showed that there may be challenges associated with the most basic step of interpreting lab results: detecting abnormal results. Many participants in this study overlooked flagged values and other abnormal results. Two subjects even falsely identified results as abnormal when they were not. We conclude, then, that these displays are failing to make abnormal results obvious and easy to detect. Developers must use evidence-based design principles to help users efficiently and accurately detect abnormal values in their online lab results and other medical reports.

This study only focused on the most basic sub-task of interpreting lab data: identifying or recognizing abnormal results. However, interpreting and using lab results is much more complicated. Unfortunately, research has shown that few online portals provide context-senstive information to faciliate interpretion of results [16]. Hence, many people subsequently turn to the Internet to aid with interpretation [17]. This contrasts general recommendations for consumer health information systems: information should be contextualized, comprehensible, and personalized to the individual [18]. Moreover, factors such as as usability, accessibility [19], and demands on eHealth literacy [20] are also essential considerations for online lab results as they are consumer health information systems and not static isolated displays.

Participant comments in our study align with findings from other consumer informatics studies; patients struggle with medical jargon and have difficulty interpreting their medical results. There are principles for effectively communicating lab results to citizens [11], but these appear to be largely ignored in current online lab portals. It is undisputed that citzens should be able to recognize abnormal results when furnished to them. This will continue to be true with advancing deployment of consumer facing technologies (e.g., OpenNotes). Moreover, published and anecdotal evidence suggests that clinicians also struggle reviewing automated and electronic laboratory results; displays are often poorly designed and require considerable mental effort to read [21]. Significant changes are required to the displays and information contained in online lab results to transform them into valuable tools for people to understand their health status and manage their conditions. Some of these changes may be appropriate for health information systems used by clinicians as well.

References

[1] B.J. Zikmund-Fisher, N.L. Exe, and H.O. Witteman, Numeracy and literacy independently predict patients' ability to identify out-of-range test results, *J. Med. Internet Res.* **16** (2014) e187. doi:10.2196/jmir.3241.

[2] B.J. Zikmund-Fisher, A.M. Scherer, H.O. Witteman, J.B. Solomon, N.L. Exe, B.A. Tarini, and A. Fagerlin, Graphics help patients distinguish between urgent and non-urgent deviations in laboratory test results., *J. Am. Med. Inform. Assoc. JAMIA*. **24** (2017) 520–528. doi:10.1093/jamia/ocw169.

[3] C. Ware, Information visualization: perception for design, 3rd ed., Elsevier/MK, Amsterdam;Boston;, 2013.

[4] Estimated Glomerular Filtration Rate (eGFR), *Natl. Kidney Found.* (2015). https://www.kidney.org/atoz/content/gfr (accessed April 14, 2021).

[5] A. Mack, Inattentional Blindness: Looking Without Seeing, *Curr. Dir. Psychol. Sci.* **12** (2003) 180–184. doi:10.1111/1467-8721.01256.

[6] R.F. Haines, A Breakdown in Simultaneous Information Processing, in: G. Obrecht, and L.W. Stark (Eds.), Presbyopia Res., Springer US, Boston, MA, 1991: pp. 171–175. doi:10.1007/978-1-4757-2131-7_17.

[7] D.J. Simons, and C.F. Chabris, Gorillas in Our Midst: Sustained Inattentional Blindness for Dynamic Events, *Perception*. **28** (1999) 1059–1074. doi:10.1068/p281059.

[8] Interaction Design Foundation, Preattentive Visual Properties and How to Use Them in Information Visualization, *Interact. Des. Found.* (n.d.). https://www.interaction-design.org/literature/article/preattentive-visual-properties-and-how-to-use-them-in-information-visualization (accessed April 13, 2021).

[9] A. Treisman, Preattentive processing in vision, *Comput. Vis. Graph. Image Process.* **31** (1985) 156–177. doi:10.1016/S0734-189X(85)80004-9.

[10] A.M. Treisman, and G. Gelade, A feature-integration theory of attention, *Cognit. Psychol.* **12** (1980) 97–136. doi:10.1016/0010-0285(80)90005-5.

[11] H.O. Witteman, and B.J. Zikmund-Fisher, Communicating laboratory results to patients and families, *Clin. Chem. Lab. Med.* **57** (2019) 359–364. doi:10.1515/cclm-2018-0634.

[12] D.R. Mutasingwa, H. Ge, and R.E.G. Upshur, How applicable are clinical practice guidelines to elderly patients with comorbidities?, *Can. Fam. Physician*. **57** (2011) e253–e262.

[13] J. Taha, J. Sharit, and S.J. Czaja, The impact of numeracy ability and technology skills on older adults' performance of health management tasks using a patient portal., *J. Appl. Gerontol. Off. J. South. Gerontol. Soc.* **33** (2014) 416–436. doi:10.1177/0733464812447283.

[14] D.R. Royall, J. Cordes, and M. Polk, Executive Control and the Comprehension of Medical Information by Elderly Retirees, *Exp. Aging Res.* **23** (1997) 301–313. doi:10.1080/03610739708254033.

[15] A. Sabahi, L. Ahmadian, and M. Mirzaee, Communicating laboratory results through a Web site: Patients' priorities and viewpoints, *J. Clin. Lab. Anal.* **32** (2018). doi:10.1002/jcla.22422.

[16] G. Mák, H. Smith Fowler, C. Leaver, S. Hagens, and J. Zelmer, The Effects of Web-Based Patient Access to Laboratory Results in British Columbia: A Patient Survey on Comprehension and Anxiety, *J. Med. Internet Res.* **17** (2015). doi:10.2196/jmir.4350.

[17] T.D. Giardina, J. Baldwin, D.T. Nystrom, D.F. Sittig, and H. Singh, Patient perceptions of receiving test results via online portals: a mixed-methods study, *J. Am. Med. Inform. Assoc.* **25** (2018) 440–446. doi:10.1093/jamia/ocx140.

[18] L. Alpay, J. Verhoef, B. Xie, D. Te'eni, and J.H.M. Zwetsloot-Schonk, Current Challenge in Consumer Health Informatics: Bridging the Gap between Access to Information and Information Understanding, *Biomed. Inform. Insights*. **2** (2009) BII.S2223. doi:10.4137/BII.S2223.

[19] L. Goldberg, B. Lide, S. Lowry, H.A. Massett, T. O'Connell, J. Preece, W. Quesenbery, and B. Shneiderman, Usability and Accessibility in Consumer Health Informatics: Current Trends and Future Challenges, *Am. J. Prev. Med.* **40** (2011) S187–S197. doi:10.1016/j.amepre.2011.01.009.

[20] H. Monkman, and A.W. Kushniruk, The Consumer Health Information System Adoption Model, *Stud. Health Technol. Inform.* **218** (2015) 26–31.

[21] B.J. Lesselroth, and D.S. Pieczkiewicz, Data visualization strategies for the electronic health record, Nova Science Publisher's, Inc, Hauppauge, N.Y, 2011.

Address for correspondence

Helen Monkman monkman@uvic.ca