



Pacific Knowledge Systems

Auto-Validation with RippleDown

Biological auto-validation of laboratory results is important in detecting possible errors in laboratory results. RippleDown can provide very high levels of biological auto-validation as a by-product of providing detailed interpretative comments of laboratory data.

Paul Compton

University of New South Wales

2013

Contents

Executive Summary.....	3
Introduction	4
Auto-Validation Techniques.....	5
Simple Techniques.....	5
Biological Validation	6
Biological Auto-Validation Systems	7
Performance of Biological Auto-Validation Systems.....	8
RippleDown Auto-Validation	9
Differences between RippleDown and other Biological Auto-Validation	10
Data on RippleDown Auto-validation.....	12
Auto-Validation Rates	12
Rejection Rates.....	13
Commencing auto-validation.....	15
Data Summary.....	18
Comparisons	19
Detailed Comparison	19
Summary	21
References	22

Executive Summary

Diagnostic laboratories have various processes in place to try to ensure that patient results sent out to referring clinicians do not contain any errors. These processes commonly given the name validation¹, include verification, confirmation and authorization. Verification is intended to provide objective evidence, that the specified laboratory requirements have been fulfilled. Confirmation of the results involves technical criteria such as quality control, limit values, delta checks and various additional checks for example on dilution, calculation etc. The last step, authorization of the result, is the acceptance of a result by a qualified person and may include technical as well as a medical checking of the plausibility of the result. In the following we will use the common term validation¹ while focusing on confirmation and authorization. A key goal for laboratories is to automate as much of this validation as possible. As Sepulveda and Young note: “Auto-validation avoids human intervention in the certification of laboratory results and is a major driver of efficiency improvements in laboratory operations” (Sepulveda and Young 2012).

In general, auto-validation techniques aim at directly detecting errors. Is the value measured for some analyte, biologically possible? Is the difference from the last measurement for that patient above normal biological variation (delta check)? If more than one analyte is measured, is the pattern of results across the analytes consistent what is known about the patient? Auto-validation levels for single results can be very high, but for more complex reports, published data suggests only 50%-70% of laboratory reports can be auto-validated and sent out without human inspection.

Auto-validation with RippleDown is approached differently. Rather than identifying whether a given pattern of results is anomalous or not, RippleDown is primarily intended to provide the referring clinician with specific clinical advice about the pattern of results for that patient. In this context, anomalous results which need checking, are patterns which are either not recognised by the system, or where the pathologist adding rules believes the results should be checked before being sent out.

In a study of over 7.5 million reports processed by 185 RippleDown knowledge bases in 12 laboratories, between April and September 2011, the overall auto-validation rate across all 7.5 million reports was 85.4%. Most other studies have looked at a much smaller number of reports in specific laboratory domains, whereas the 185 knowledge bases studied here cover a very wide range of chemical pathology. Although RippleDown is used in later stage of the workflow than much other validation 85.4% is an extremely high rate of overall auto-validation. At the same time highly specific advice is being provided by the pathologist to assist in the interpretation of the results, whereas for conventional auto-validation approaches, clinical interpretation is a further task, beyond validation. Many authors have pointed out that validation of laboratory results is only one part of reducing errors in the usage of laboratory data, and Plebani in particular has proposed that interpretative reporting is critical to reducing post-analytic errors (Plebani 2009). RippleDown then fulfills a dual role in auto-validation by reducing the possibility of erroneous results being sent out, and further assists in helping to ensure results are used appropriately, by providing clinical advice.

¹ In ISO 15189 this term is also used in another context: “confirmation through the provision of objective evidence that the requirements for a specific intended use, application or test have been fulfilled”.

Introduction

The goal of laboratory medicine is to provide diagnostically relevant and accurate information about a patient. The literature suggests that errors related to laboratory testing range from 0.1%-9.3% (Kalra 2004). These errors can occur in the pre-analytic phase: for example the quality of the sample collected, inappropriate test ordering, transcription; during analysis: equipment or reagent malfunction or miss-calibration; or post-analysis: the report reaching the right person, information relevant to the patient's management being noted etc. An Australian study indicated that the median error rate simply for transcribing requests was between 1% and 3% with one laboratory having a 13% error rate (Khoury, Burnett et al. 1996). Although one might argue that a clinician's failure to note important information on a report is outside of a laboratory's responsibility, Plebani argues that using tools like interpretative reporting, laboratories can try to assist in ensuring laboratory results have the appropriate impact on clinical care (Plebani 2009). Of the errors that occur, reports suggest that 46%-68% of errors are pre-analytic, 7%-13% are analytical errors, and 19%-47% are post-analytic (Bonini, Plebani et al. 2002; Plebani 2006).

Pacific Knowledge Systems provides its RippleDown Data Entry Auditor software to assist in reducing pre-analytic error. It is widely used in checking that appropriate test have been ordered given the patient's history, previous test order, insurance payment regulations etc. However, many pre-analytic and post-analytic errors need to be addressed primarily by quality assurance rather than to automatic checking. For example whether the correct patient identification label has been applied to a specimen can only be addressed by appropriate quality assurance processes, unless there are previous results or other information about the patient, that are not consistent with the results from the mislabeled sample.

There are a number of aspects of quality assurance techniques for analytic error which can lead to concern. The first issue is the quality of a set of results that have been processed where inter-assay error needs to be considered. Laboratories routinely use quality control samples in each assay to assess inter-assay changes and techniques have long been available on how to use the data from QC samples (Levey and Jennings 1950; Westgard, Barry et al. 1981). There are also quality assurance issues in how results may be compared between laboratories, particularly in respect to the reference intervals that might be used (Sikaris 2012).

The next issue is detecting potential errors in individual results. Some analysers directly provide information on whether an individual sample has been measured correctly. This is normally confined to issues such as whether there has been sufficient sample or whether it is haemolysed etc. However, there are many other causes of errors and the next stage of validation is to check whether results make sense. This is the area where auto-validation, automatic checking that results make sense, is applied.

Auto-Validation Techniques

Simple Techniques

Simple auto-validation techniques (often called auto-verification) are expected to be in use in every laboratory (CLSI 2006; CAP 2009). They include limit checks: checking whether an individual result is outside the range that is biologically likely, or delta checks: whether the change between measurements on two consecutive occasions is biologically feasible. Despite many papers confirming the desirability of such systems, there is surprisingly little detailed information about their performance. Rogoski comments that middleware enables auto-verification of 80%-90%, while LIS based auto-verification only reaches about 50%, but this is essentially a magazine article so no references are provided (Rogoski 2010).

A recent evaluation of delta and limit checks on over 800,000 patient results in one laboratory (Shih, Chang et al. 2011) found the following: Although 95.6% of individual results in this particular study could be auto-verified, this dropped to 81.5% when request verification was considered, i.e. verification of all the results requested. When looking at individual tests some had auto-verification levels as low 70% or less.

Recommendations on the use of auto-verification also suggest using rules to examine whether patterns of results for a patient are consistent. Shih et al comment that it was difficult to write such rules and in their study only a few types of results were checked for consistency. They apparently used about 30 pattern rules; however, detailed results are not reported in the paper and the results essentially concern delta and limit checks. There is no discussion in the paper of whether results were appropriately auto-validated.

Dorizzi et al. report on a simple rule based system for an endocrine an tumour marker lab, with rules for both delta checks and patterns of results (Dorizzi, Caruso et al. 2006). Dorizzi et al. developed 13 auto-verification rules and in a six month trial 80% or results were auto-verified. For example the thyroid rules are:

```
If TSH <0.35 mU/L And FT4 > 23 pmol/L Then AUTO-VALIDATION  
If TSH <0.35 mU/L And FT4 <23 pmol/L No AUTO-VALIDATION; If FT3 >  
5.7 pmol/L AUTO-VALIDATION  
If TSH > 4.3 mU/L And TPO > 200 KU/L AUTO-VALIDATION
```

A comparisons of these results with a similar RippleDown knowledge base is included below.

Biological Validation

Biological validation is distinguished from “scientific” or “technical” validation, and considers whether results are biologically plausible. This can include population considerations, but the more common meaning relates to individual patients. That is, that any laboratory reports that are sent out by the laboratory are consistent with what is known of the patient: the clinical and results history and any other relevant information. Simple limit and delta checks are generally seen as part of scientific validation, but also contribute to the concept of biological validation. E.g. A delta check, generally seen as a stand alone test, is really part of checking that the changes over time are consistent with all information available about the patient.

A major issue with the notion of biological validation is that there is a wide range of interpretation of what results might be consistent with patient information and can be sent out without further consideration. In two studies evaluating Valab, the laboratory auto-validation expert system (Valdiguie, Rogari et al. 1992), laboratory experts were asked to inspect the same reports processed by Valab and advise whether or not they should have been auto-validated. In both studies the auto-validation levels recommended by the experts ranged from 75% to 95% (Valdiguie, Rogari et al. 1992; De Smet, Nanos et al. 2009). Clearly experts are not simply identifying possible errors as one would assume that this would lead to a higher level of agreement, but presumably are making broader decisions about what sorts of results should be manually reviewed. We presume this relates primarily to giving a pathologist the opportunity to provide advice to the referring clinician, whether by adding a comment to the report, or phoning the clinician etc. One of the Valab evaluations was based on 591 haematology reports – all the reports in a day (De Smet, Nanos et al. 2009). The consensus of seven experts was that 82% could be auto-validated (with the experts ranging from 75% to 95%). In contrast in another study using different auto-validation technology, 95% of full blood counts were auto-validated (Burgess, Robin et al. 2009). It is not likely that there was a such a high error rate in De Smet’s laboratory compared to Burgess’s, rather it seems likely that there are different policies in play about when interpretative advice is needed.

Biological Auto-Validation Systems

Although it is clear that biological validation includes the broader concern of giving experts an opportunity to provide advice to referring clinicians, conventional auto-validation software seems to be based around the concept of validating individual laboratory results one by one – checking whether they are anomalous in the light of other data.

The following example rules are all taken from Valab

If there is a request for serum potassium,
And there is no preceding result,
And the potassium is abnormal (>5.8 mmol/L),
And the plausibility criteria (high creatinine, high urea, high uric acid, low calcium, specimen coming from nephrology ward or intensive-care unit) are present,
Then this abnormal potassium result is accepted
(Valdiguie, Rogari et al. 1992)

If there is a decrease of serum calcium with low serum calcium.
And there is a result and a previous result for creatinine.
And the creatininemia is above 300 $\mu\text{mol/l}$.
And the value of creatininemia increases.
And this increase is above 100 $\mu\text{mol/l}$.
Then increase the delta-check of calcium by 10%.
(Prost and Rogari 2002)

If the PT/QT increases by more than 8 s.
And there is a result of activated partial thromboplastin time (APPT).
And there is an increase of APPT of less than 3 s.
Then it is not possible to validate such a value of PT/QT.
(Prost and Rogari 2002)

Such an auto-validation system requires rules that cover each legitimate data profile. If the data matches any of these rules, the result will be auto-validated. If no such rule applies the case is referred for manual validation by a human expert. There can also be rules that will specifically refer a data pattern for manual validation, as shown in the last example above. Rules can have specific numerical values embedded or refer to standard reference range and standard delta values for changes between repeat measurements and so on.

To build a set of rules for this sort of approach one needs to think of all the different patterns of data and information which might relate to a given analyte result. This is a difficult task. On the other hand there is a risk that the rules constructed are not sufficiently specialised, and reports containing anomalous data may be auto-validated, as will be discussed further below.

Performance of Biological Auto-Validation Systems

Guidi et al. document auto-validation of 92% for urgent stat tests using Valab. Of the results that were not auto-validated, 93% were released after manual inspection and only 7% of the manually validated results led to a retest or a phone call to the referring clinician (Guidi, Poli et al. 2009). These are essentially similar to the results reported above (Shih, Chang et al. 2011) and presumably given that these are stat tests, these results come essentially from delta and limit check.

In Prost's more wide-ranging study of Valab the mean rate of auto-validation in a general hospital was around 70% of but varied between 50% and 90% depending on various factors (Prost and Rogari 2002). For example a report with less than 3 results is validated 85% of time but a report with between 10 and 21 results it is validated only 59% of the time.

Burgess et al. describe a system where haematology rules which the laboratory had developed over 25 years were implemented using the Cerner Discern Expert Module (Burgess, Robin et al. 2009). The system auto-verifies 95% of full blood counts while 5% are referred for operator intervention and verification. No data was provided on the accuracy of the rules used. A key feature of this result is that auto-validation rules have been developed over many years, by the experts whose role it is to validate haematology reports. These results show the level of auto-validation that can be achieved when a close match is achieved between the rules that have been assembled and the validation practice of the experts. Given the availability of the Discern Expert Module and its integration with the Cerner LIS it is surprising that there are not other reports of rule-based auto-validation using Cerner or other similar systems. The most studied rule-based auto-validation system is Valab which commenced in 1988.

In a 2009 review Valab automatically validated 56.5% of 591 haematology reports (De Smet et al. 2009). Laboratory experts were asked to advise on which of the 591 reports should have been auto-validated. The consensus of the experts was that 86.1% of results could have been automatically validated. On the other hand 29.3% of the 13.9% of results that experts said should not be validated, were auto-validated by Valab. These results are remarkably similar to the initial evaluation of Valab (Validguie et al 1992). In the original study Valab auto-validated 56.5% of 200 reports which had abnormal results (as identified by the LIS). The consensus of the experts was that 82% of the reports should have been auto-validated. The system also auto-validated 25% of the 18% of results that experts said should not be auto-validated. According to the report from 1988 to 1992 Valab has processed about 70% of the total laboratory workload and auto-validated >50% of reports. A review of LabRespond, a broadly similar system to Valab, showed similar performance (Oosterhuis, Ulenkate et al. 2000).

RippleDown Auto-Validation

RippleDown takes a quite different approach to auto-validation. In Valab-type validation the purpose of the rules is to make the decision about whether a particular individual result can be auto-validated. If a rule is satisfied and there is no overriding rule requiring manual validation, then the result is auto-validated. With RippleDown, rules are created to add interpretative comments to a report. There is then a secondary decision, when the expert decides that a particular report with perhaps a sequence of comments can be auto-validated. This decision is not made when the rules providing these comments are built but after observing these rules in action for some time.

Statistics are provided on how often a comment has been given without being changed, and the pathologist can use these statistics to make the judgement, that a certain report for a particular profile of results has been given correctly so often, that it can be safely auto-validated. This is exactly the same type of judgement that is made in every aspect of laboratory medicine: that a system or method has been tested sufficiently thoroughly that it can be put into use. A pathologist can use these statistics to make the judgement, that a certain report for a particular profile of results has been given correctly so often, that it can be safely auto-validated. This is exactly the same type of judgement that is made in every aspect of laboratory medicine: that a system or method has been tested sufficiently thoroughly that it can be put into use.

Experts do not simply turn on auto-validation, they set a percentage auto-validation level for that type of report. This percentage can be changed over time as appropriate. RippleDown will then randomly select cases to refer for manual validation according to the percentage auto-validation for that type of report. This allows the on-going performance of the system to have some level of checking.

In fact, using RippleDown experts can choose to set different levels of auto-validation for different sections of the report which are generated by different sets of rules. That is, a "Recommendation" section of the report may have a lower auto-validation level than a "For further information" section.

Differences between RippleDown and other Biological Auto-Validation

Both RippleDown and Valab rules (for example) identify particular data patterns and are likely to have the same type of rule conditions, but they have different types of conclusion. RippleDown is generally used to provide advice about the pattern of results, related to the patient's clinical history and other information. This advice is intended to assist in the patient's management by their clinician and perhaps advise further investigation. In contrast Valab rules simply advise whether or not the results can be released or needs manual checking. Valab has essentially two conclusions for rules, whereas RippleDown allows as many conclusions as the expert wishes.

For Valab the decision to auto-validate is the actual conclusion of the rule – auto or manual validation are the only two decisions the system makes. For RippleDown, most of the rules are used to identify the type of pattern the data represents and to generate the interpretive report, and then there is a decision on top of this whether this type of pattern should be auto-validated and at what level. This decision is based on the expert's cumulative experience in seeing the rule in action. RippleDown also allows the expert to construct Valab type rules, which explicitly refer certain types of case for manual validation.

For systems like Valab, rules are constructed for specific analytes, so a set of rules will be accumulated for all the patterns relevant to a specific analyte such that you can say the system auto-validates analyte X. There is the underlying problem that the rules covering analyte X may not (and probably cannot) be complete and correct, but the intention of the knowledge base developer is to completely cover the patterns for X. In contrast RippleDown provides auto-validation for domains rather than specific analytes: thyroids, lipids, diabetes etc, with a knowledge base for each domain. The choice of how to structure domains is left to the pathologist, they could have one comprehensive knowledge base for all of biochemistry or they may construct many smaller knowledge bases. In creating a RippleDown knowledge base, the expert specifies which selection of results, if ordered by the referring physician, will be processed by a particular knowledge base. All results available, plus all the other information specified as relevant (e.g. the last 10 results, the referring clinician's notes etc) are also passed to the knowledge base and can be used in generating the report. The same set of patient data might be passed to a number of different knowledge bases, each generating a report for its specific area of "expertise". These reports may then be combined as sections of a single and comprehensive interpretive report. The final report may even have a summary section generated by a knowledge base that picks out the relevant points from the other sections..

The simplest way to compare the scope of RippleDown validation to validation based on analytes, is to say that RippleDown can provide auto-validation for all of the analytes relevant to a particular knowledge base; i.e. a particular diagnostic domain.

With conventional auto-validation, the only data routinely available on the system's performance is the auto-validation rate. There is no information on what happens to manually validated results unless a study is specifically carried out like those described above. RippleDown's primary role is to provide interpretative reports to referring clinicians. Manual validation may result in the report being sent out with the RippleDown report unchanged, or else the report is corrected. This means statistics on the "rejection" of RippleDown reports are available. This data is sometimes difficult to interpret, because the decision that is recorded is not whether the advice on the report represented

a correct clinical assessment of the data and clinical information available, but whether the advice was sent to the referring clinician as is.

Some specific comments, although “correct” might not be appropriate to be sent out for a particular patient of a particular clinician. Or the validator responsible for the manual validation may be different from the expert who builds the rules for the domain, and the validator may wish to make some very some very small changes in the wording of a comment. Anecdotal reports suggest that senior pathologist trainees, may wish to put their own stamp on the comments sent out, and identify small differences in the data from the normal pattern of results for that comment. They may or may not refer the case for subsequent rule-building - an option is provided to send the case for rule building when a report is rejected (including any minor editing) On the other hand if they do refer the case to the rule builder for a change in rules to produce a new report, the more senior expert responsible for rules building, may decide that, in fact, the original wording is more appropriate.

Another scenario which distorts the “rejection” rate is where certain comments are added as “internal” comments for the validator staff, e.g. to check a suspect test result, or to add a further test to the order (if RippleDown’s reflex testing capability is not being used). These internal comments are meant to be acted upon by the validator user, then removed from the report before releasing to the referring physician. These are recorded in the statistics as “rejected” reports, but there is no error in the report.

Because statistics on report rejection only measure whether or not the comment generated was sent out as is, and include many cases where the conclusion was not in anyway erroneous, they do not provide a useful measure of actual errors in comments. The best measure of genuine errors in report comments may be the rate at which rules are built. This represents the rate at which the senior expert decides a report for a particular type of data pattern should be changed. This may be a minor refinement or a more significant change, but represents the view of the senior expert responsible for rule building that different advice should be given in the future for that given pattern of results.

Data on RippleDown Auto-validation

The data below come from monthly logs of the number of reports processed by each RippleDown knowledge base, the number auto-validated and the number of manually validated reports which are “rejected” and the number of rules developed.

Auto-Validation Rates

From April to May 2011, 12 laboratories auto-validated 7.5 million (85.4%) reports and manually validated 1.1 million (14.6%) reports. The behaviour of the laboratories varies considerably with auto-validation rates varying from 48% to 96% except for one laboratory that manually validates all the results for the single knowledge base it runs. The range of auto-validation for different laboratories is shown in Figure 1. Figure 2 shows the numbers of knowledge bases at different levels of auto-validation across all laboratories. In total there are 185 knowledge bases across 12 laboratories. As can be seen from Figure 2, 81% of reports are processed by knowledge bases where are over 80% auto-validated. It should be noted that these auto-validation rates are the overall rates for each knowledge bases and each report in the knowledge base may have a different level of auto-validation.

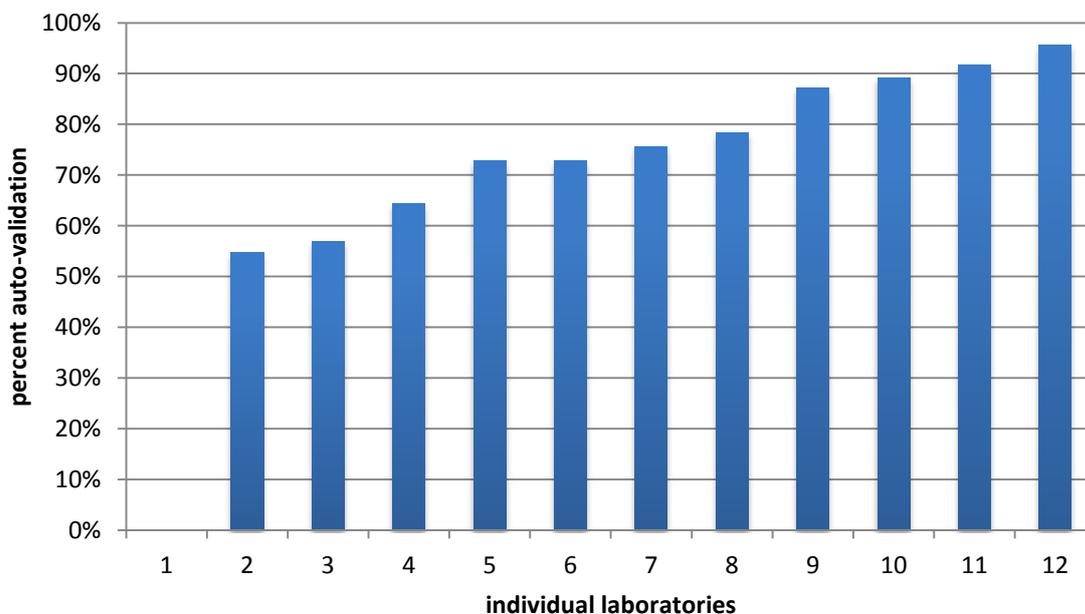


Fig 1. Auto-validation rates for different laboratories. The total auto-validation rate for all reports from all of a laboratory's knowledge bases is shown

Although there are a large number of knowledge bases where all reports are manually validated, they process only about 4% of the cases. These are specialist knowledge bases where the volumes are so low there is little to be gained by auto-validating. One laboratory which runs a single knowledge base interpreting bone-mineral density results, manually validates all its reports and also processes the smallest number of reports. This laboratory is also interesting in that 2.5% of reports were rejected, and rules were constructed for 53% of the rejected reports, much higher than the 3.5% across all laboratories, as discussed below. This laboratory uses RippleDown to write detailed letters to clinicians, explaining the important features of a patient's bone mineral density results and recommended management advice taking into consideration significant clinical history including

previous results, fractures, height and weight changes, medications, calcium intake and so on. The large majority of reports, 97.5%, are accepted as appropriate, but it seems likely that the comparatively high rate of writing rules is to keep improving the quality of the letters in this highly complex domain

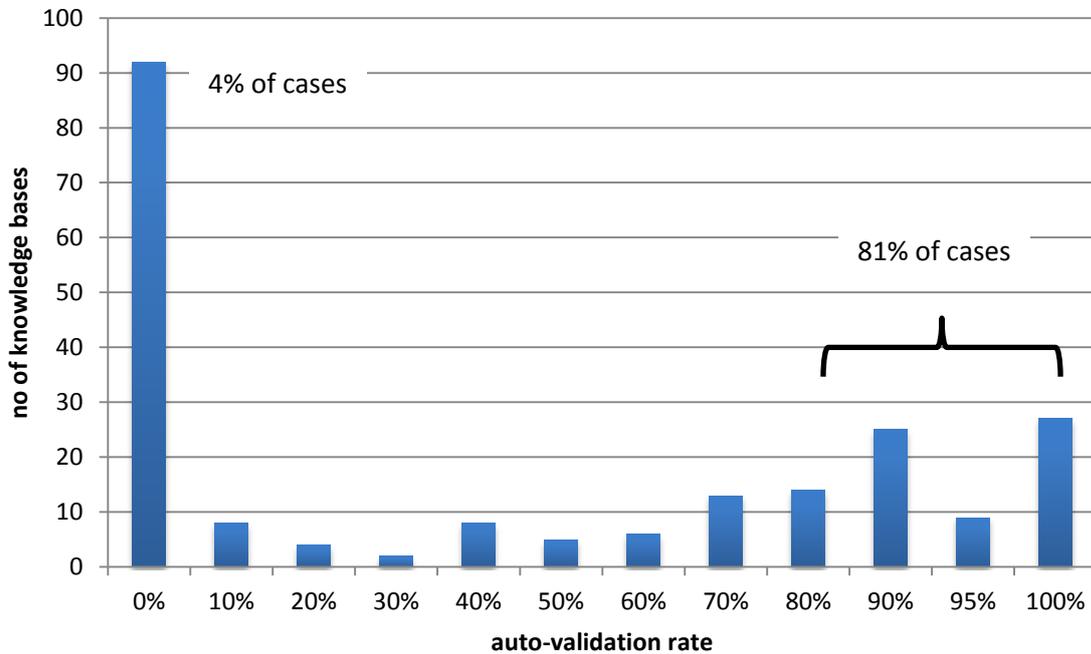


Fig 2 The number of knowledge bases at each level of auto-validation. 4% of all the cases were processed by knowledge bases where all validation is manual, and 81% of cases were processed by knowledge bases with 80% or higher auto-validation. The data comes logs of use from April to September 2011

It is clear that laboratories adopt domain specific policies with respect to choosing between manual and auto-validation. Figure 3 shows the number of knowledge bases for each laboratory that are partly (or totally) auto-validated or entirely manually validated. Clearly laboratories adopt domain specific policies about auto-validation, as most laboratories have both partly auto-validated and entirely manually validated knowledge bases.

Rejection Rates

As discussed previously, reports are “rejected”, not only because the advice may be incorrect or not sufficiently targeted to the patient, but because part of the comment may be for the validator themselves and is not intended to be sent out, or the validator has made some minor refinement to the interpretation. In some instances either the validator or the knowledge builder has decided it is not appropriate to build rules to cover such differences. Therefore, as mentioned previously, the rejection rate is best measured by the rate of rule addition. This is only 3.5% of the rejected reports which corresponds to just 0.3% of the manually validated reports.

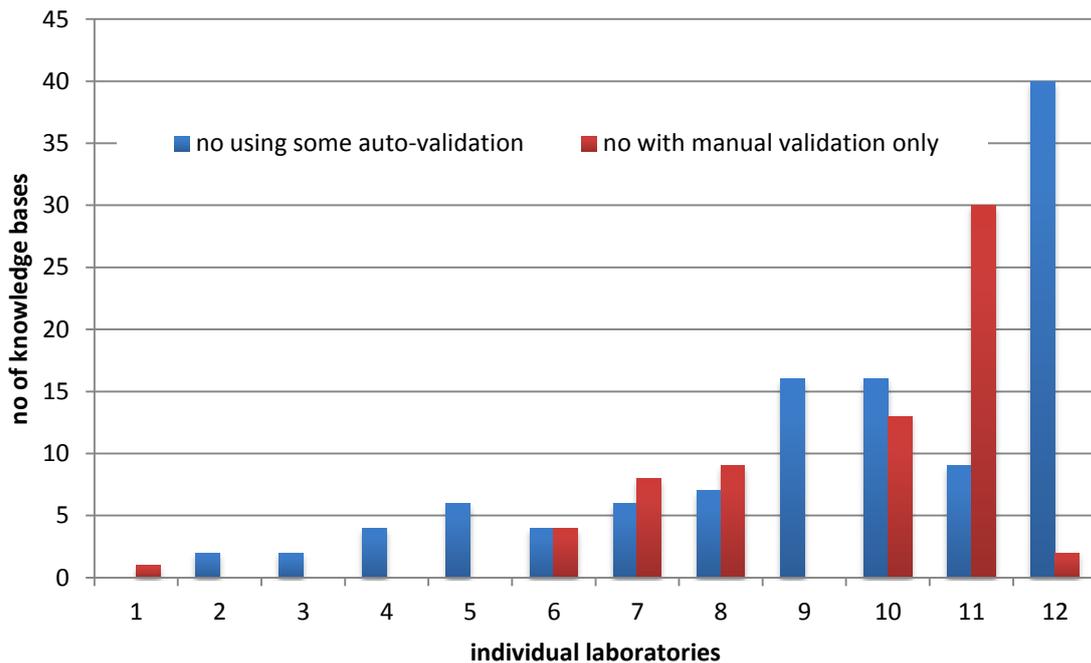


Fig 3. The number of knowledge bases for each individual laboratory which are partly (or completely) auto-validated or entirely manually validated

We can first note that the rejection rate for knowledge bases that are entirely manually validated is very similar to the rejection rate for knowledge bases that are only partly manually validated. The overall rejection rate is 9.7% while the rejection rate for entirely manually validated knowledge bases is 10.6%. Perhaps this reflects the level at which validators (particularly senior pathology trainees) are willing to spend time making minor changes to reports which do not warrant a new rule being added. Given that 85.4% of reports are auto-validated, the overall rejection or edit rate is about 1.5%.

Further details of the rejection rates are shown in figures 4 and 5. Figure 4 shows that rejection rates within laboratories are independent of the level of auto-validation. Figure 5 shows the numbers of laboratories with different rates of report rejection. Clearly different practices are at play because of the wide range of rejection for different knowledge bases. The majority of reports are processed by knowledge bases with rejection rates of between 1% and 10%, with the overall average being 9.7%.

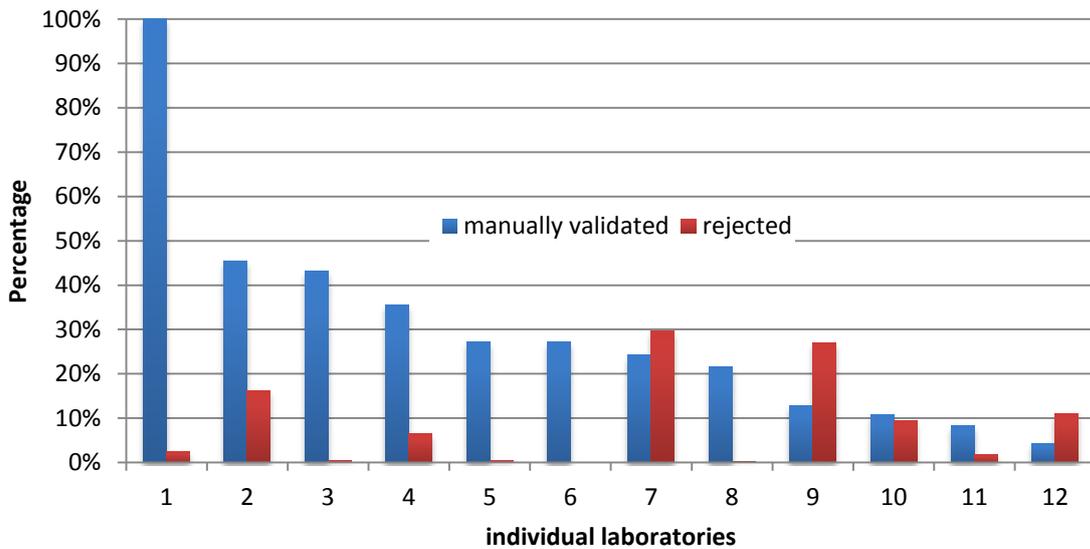


Figure 4: the rejection rate of manually validated reports for various laboratories is shown together with the rate of manual validation

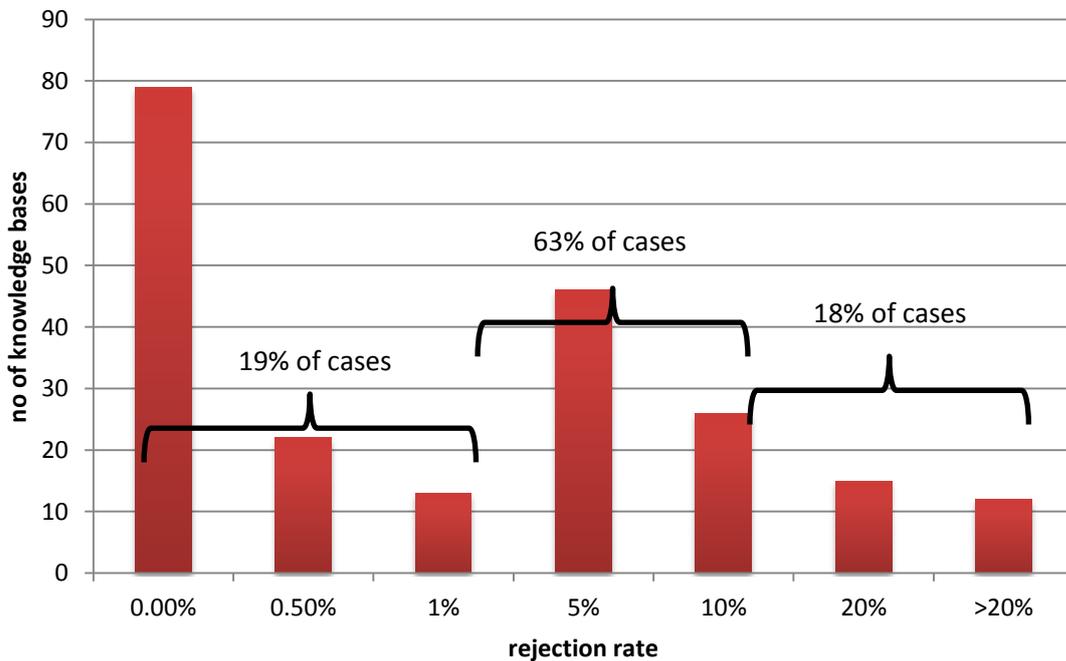


Fig 5. The number of knowledge bases with different rates of report rejection. The proportion of the cases processed by knowledge bases with different rates of rejection is also shown

Commencing auto-validation

As a laboratory moves from manual validation to some level of auto-validation, the rate of rejection of manually validated cases is likely to increase. Laboratories will only choose to auto-validate

reports where they have considerable confidence that the particular report is always given appropriately. This means that reports likely to be rejected i.e. changed in some way, will be concentrated in the manually validated reports. This means that as a knowledge base moves from manual validation to some level of auto-validation there is likely to be a relative increase in rate of rejection of the remaining manually validated cases. Offsetting this effect however, is the fact that as the rejected cases are referred for rule building, then rules to provide the correct reports will be added so the same type of cases will not be rejected in the future, thus tending to lower the rejection rate. With both factors in play interpretation of the data is difficult.



Fig 6. The blue line shows the auto-validation rate for one particular knowledge base used to provide interpretative comments for THA reports. Each point corresponds to the log data for one month. The red line shows the rate at which manually validated reports were rejected. The number of rules added in relation to each of the two increases in rejection rates are shown. Each data point represents one month.

Figure 6 shows an example of a single knowledge base moving from manual validation only to various levels of auto-validation. Auto-validation increases in three steps from 0% to 14%, then to 40%-50% and then to 80%. The X axis shows the number of cases processed, with each data point corresponding to one month's activity. The decisions to increase auto-validation are based on significant experience with the system's performance. Validation is increases to about 14% after about 2000 cases have been seen, to 40% after 4,000 cases and to 80% after 8,000 cases have been seen. Again these are composite figures combining different levels of auto-validation for different reports across all cases processed. Obviously very common reports are likely to be set at a high level of auto-validation much earlier than reports for more unusual profiles.

The initial rejection level for this knowledge base is about 7%, but after 7 rules are written the rejection rate drops close to 0. When auto-validation increases to about 50% the rejection rate

increases and a further 43 rules are written leading to auto-validation increasing to about 80%. It is interesting to note how the rejection rate drops to zero without further rules being added. This suggests that the staff validating the reports have decided that the small changes they had been making were not really warranted.

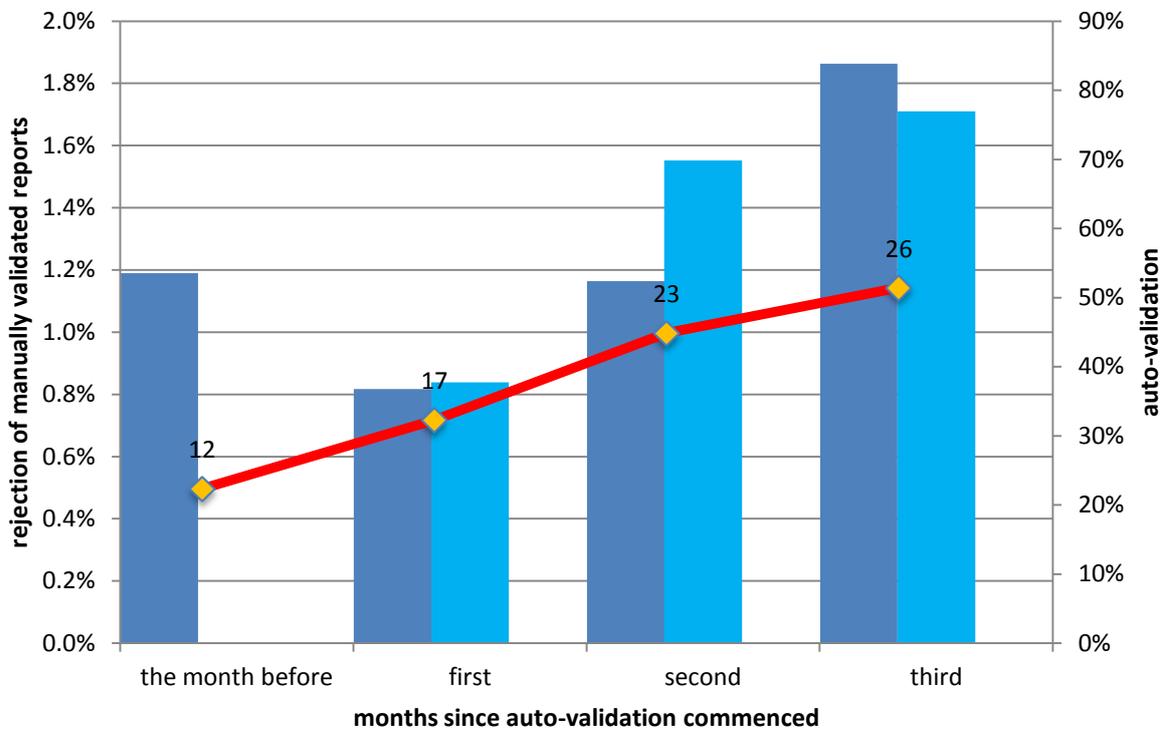


Fig 7: This figure shows data around the commencement of auto-validation for the 23 knowledge bases for which this data is available. The purple columns show the percentage of results that were auto-validated each month, starting the month before auto-validation was turned on. The blue columns show the percentage of manually validated reports that were rejected. The red line shows the cumulative number of rules added per knowledge base averaged across 23 knowledge bases.

Figure 7 shows composite data for 23 knowledge bases where the monthly log data shows an initial period of manual validation only². Data is shown for the month before auto-validation commences and then for the next three months after auto-validation commences. In the first month the auto-validation rate is 38% increasing to 70% and then 77% in the second and third months. The rules shown are the cumulative number of rules added per knowledge base over the 4 months. The rejection rate initially drops, presumably because of the rules added, but then increases each month. The increase in rejection rate does not indicate more cases are rejected, but that the rejected cases have been concentrated in the decreasing number of manually validated cases. In effect, the knowledge base has flagged those reports that are at more risk of needing correction prior to release.

² The volume of cases processed by most knowledge bases is so high that the monthly log from the first month's use show some auto-validation already being used

Data Summary

RippleDown provides for a very high level of auto-validation with overall 85.4% of reports being auto-validated. The level of auto-validation varies between laboratories with levels ranging from 48% to 95% across laboratories except for one specialist laboratory with a single knowledge base with 0% auto-validation. There are a large number of knowledge bases which have no auto-validation and all reports are manually validated; however, these are for very low volume sub-domains. The bulk of all reports, 81%, are processed by knowledge bases which show an average level of 80% auto-validation. Of all the reports that are manually validated, about 10% of them are changed before release. That is, of all the reports that are generated including manually validated and auto-validated, about 1.5% are changed before being released.

This leads to the conclusion that in 98.5% of cases, the report provided by RippleDown systems provides the appropriate clinical advice to the referring clinician, and to achieve this, 14.6% need to be manually validated. More significant issues that warrant a change in the rules occur in 0.3% of cases. These changes may not be of great clinical significance, for example the expert may decide that another comment will be more useful for this type of case in the future.

Comparisons

Although biological validation as developed in systems like VALAB relates to RippleDown, in that both consider patterns of patient results, some caution is needed in making direct comparisons. RippleDown is used in addition to other validation measures laboratories might be using. For example a laboratory might already have software that carries out delta, limit and critical value checks, which is applied before the data is passed to RippleDown. VALAB includes these sorts of simple checks as well as looking at patterns. This means that detailed comparisons cannot be made, but the differences in performance are sufficiently large that more coarse-grained comparisons can be made. If one considers only simple checking like delta checks and limit checks then very high levels of auto-validation can be achieved e.g. about 95% in (Shih, Chang et al. 2011). On individual results, where only simple checking applies, VALAB achieved auto-validation of 92% (Guidi, Poli et al. 2009). Prost's study of Valab reports demonstrated that if they contained less than 3 results, reports were validated 85% of time whereas reports with between 10 and 21 results were validated only 59% of the time. They showed the mean rate of auto-validation of reports in a general hospital to be around 70%, with a wide variation - between 50% and 90% depending on various factors (Prost and Rogari 2002). One would assume the validation of reports with less than 3 results was dominated by single result tests where only simple checking is possible giving a high auto-validation rate. Across a whole range of laboratories and domains, with varying numbers of results RippleDown provided 85.4% auto-validation. It is also worth noting that with conventional validation the expert is still required to enter comments into the report if advice to referring clinicians is needed.

Detailed Comparison

To some extent a more detailed comparison between RippleDown and a more conventional approach can be made using the data of Dorizzi, Caruso et al. (2006). They developed 13 auto-verification rules for an endocrine and tumour marker lab. For example the thyroid rules are:

If TSH <0.35 mU/L And FT4 > 23 pmol/L Then AUTO-VALIDATION
If TSH <0.35 mU/L And FT4 <23 pmol/L No AUTO-VALIDATION; If FT3 >
5.7 pmol/L AUTO-VALIDATION
If TSH > 4.3 mU/L And TPO > 200 KU/L AUTO-VALIDATION

Using rules such as these during a six-month trial (about 100,000 reports) 80% of reports were auto-verified. PKS customers do not necessarily group hormones and tumour markers into the one knowledge base so it is difficult to make exact comparisons, but taking the 10 knowledge bases out of the 185 that relate to hormones or tumour markers, the auto-validation level was 86.7% across 931,604 cases processed during the six-month period. If one considers only thyroid knowledge bases, as presumably fairly similar, one finds that although the average auto-validation is 87.1% there is considerable difference between laboratories, ranging from 73.1% auto-validation to 98.0% auto-validation. The variation might be due to different practices of different experts, or different types of patients being referred or perhaps for labs with higher volumes there is greater need to auto-validate. The largest lab with 371,771 reports had an auto-validation level of 86.4%, while the second largest with 197,719 reports had an auto-validation level of 93.9%. Conversely, the two labs with the lowest number of reports had the smallest (73.1%) and highest (98.0%) auto-validation rates. Clearly report volume is not a the only factor resulting in different auto-validation levels.

If we take the two largest labs, with volumes exceeding the volumes in the Dorizzi study, using RippleDown resulted in decreases of 32% and 69% respectively, in the amount of manual checking required compared to the results from the Dorizzi laboratory study. Again, these comparisons are not exact because the Dorizzi rules include delta and limit checks, whereas these may have been carried out already in the labs using RippleDown.

Summary

RippleDown supports various aspects of validation and auto-validation. The RippleDown Data Entry Auditor enables laboratories to write rule to reduce errors in pre-analytic data entry, for example checking that appropriate tests have been ordered given the patient's history etc. RippleDown also enables laboratories to write rules to flag particular laboratory reports for manual inspection, because results are anomalous, or the pathologist wants to manually check particular types of reports etc. The value that RippleDown has for such validation applications is that the laboratory does not depend on a black box system, but can write and change rules to suit local requirements very rapidly. Published data across many knowledge bases show a rule can be added and the knowledge base revalidated ready for use, on average in about two minutes (Compton et al. 2011).

There is a further unique contribution that RippleDown makes to laboratory auto-validation. RippleDown enables pathologists to very write rules to provide highly patient-specific clinical interpretation of laboratory results. But in writing these rules the pathologist is also identifying very specific clinical entities. The central idea behind so-called biological validation for individual patient results is to identify if the pattern of results for a patient is biologically plausible: is it consistent with past results and any clinical information about the patient that is available. When a pathologist writes a rule to interpret some results, they take into account whatever patient information is available in the laboratory. That is, by identifying the clinical entity that the results and other information correspond to and writing a clinical comment about this clinical entity, they are also judging the biological plausibility of the results (or perhaps that the results are implausible).

RippleDown goes a step further than simply identifying results as biologically plausible or implausible. When an expert provides a rule for a knowledge base there is always the possibility that there will be some exception to the rule, which the expert didn't identify, because the exception didn't arise in that context. RippleDown provides a Validator module where reports are checked, and any exceptions or need for further rules are flagged for the expert adding rules to address. The Validator provides statistics on how often a clinical interpretation is changed, thus providing information on how reliably the clinical entity has been identified. The pathologist can then choose to auto-validate highly reliable interpretations. They can also chose some level of partial auto-validation, or no auto-validation for particular clinical entities, that they believe should be manually checked.

This means that RippleDown not only supports auto-validation based on biological validation, but enables laboratories to make a judgement on how reliable the biological validation is for each particular pattern, before auto-validation is used. With RippleDown this is not a black box, but is all under the control of the pathologist who can write rules to identify whatever patterns they think are appropriate for their particular practice and can determine the level of auto-validation that is appropriate for each clinical entity or pattern. Again, on average it takes only two minutes to add a rule, so the whole process is under the control of the pathologist as a very minor addition to their normal workflow. Data from RippleDown users across a range of laboratories and domains show that the overall level of RippleDown auto-validation is about 85%.

Finally, RippleDown contributes further by providing detailed patient specific clinical advice helping to ensure that laboratory results are more likely to be fully taken into account in patient management.

References

- Bonini, P., M. Plebani, F. Ceriotti and F. Rubboli (2002). "Errors in laboratory medicine." Clinical chemistry **48**(5): 691-698.
- Burgess, P., H. Robin, M. Langshaw, G. Kershaw, R. Pathiraja, S. Yuen, C. Coad, N. Xiros, G. Mansy and R. Coleman (2009). "Rule based processing of the CD4000, CD3200 and CD Sapphire analyser output using the Cerner Discern Expert Module." International Journal of Laboratory Hematology **31**(6): 603-614.
- CAP (2009). Lab General Checklist AUTOVERIFICATION GEN.43850~43893, College of American Pathologists.
- CLSI (2006). Autoverification of Clinical Laboratory Test Results; Approved Guideline. AUTO10-A. . Wayne, PA, Clinical and Laboratory Standards Institute.
- Compton, P., L. Peters, T. Lavers and Y.-S. Kim (2011). Experience with long-term knowledge acquisition. Proceedings of the Sixth International Conference on Knowledge Capture, KCAP 2011. Banff, Alberta, Canada, ACM: 49-56. (PKS white paper <http://www.pks.com.au/technology/white-papers>)
- De Smet, D., N. Nanos, G. A. Martens, L. Smet, S. Damiaens, G. Van Moer, A. De Pauw, K. Jochmans and M. De Waele (2009). "Automated validation of clinical laboratory test results: do we get what we expect?" Clinical Chemistry and Laboratory Medicine **47**(3): 379-380.
- Dorizzi, R. M., B. Caruso, S. Meneghelli and P. Rizzotti (2006). "The DNSev™ expert system in the auto-verification of tumour markers and hormones results." Accreditation and quality assurance **11**(6): 303-307.
- Guidi, G. C., G. Poli, A. Bassi, L. Giobelli, P. P. Benetollo and G. Lippi (2009). "Development and implementation of an automatic system for verification, validation and delivery of laboratory test results." Clinical Chemistry and Laboratory Medicine **47**(11): 1355-1360.
- Kalra, J. (2004). "Medical errors: impact on clinical laboratories and other critical areas." Clinical biochemistry **37**(12): 1052.
- Khoury, M., L. Burnett and M. A. Mackay (1996). "Error rates in Australian chemical pathology laboratories." Medical journal of Australia **165**(3): 128-129.
- Levey, S. and E. Jennings (1950). "The use of control charts in the clinical laboratory." American journal of clinical pathology **20**(11): 1059.
- Oosterhuis, W. P., H. J. L. M. Ulenkate and H. M. J. Goldschmidt (2000). "Evaluation of LabRespond, a new automated validation system for clinical laboratory test results." Clinical chemistry **46**(11): 1811-1817.
- Plebani, M. (2006). "Errors in clinical laboratories or errors in laboratory medicine?" Clinical Chemical Laboratory Medicine **44**(6): 750-759.

Plebani, M. (2009). "Interpretative commenting: A tool for improving the laboratory? clinical interface." Clinica chimica acta **404**(1): 46-51.

Prost, L. and E. Rogari (2002). "How autoverification through the expert system VALAB can make your laboratory more efficient." Accreditation and Quality Assurance: Journal for Quality, Comparability and Reliability in Chemical Measurement **7**(11): 480-487.

Sepulveda, J. L. and D. S. Young (2012). "The Ideal Laboratory Information System." Archives of Pathology & Laboratory Medicine.

Shih, M.-C., H.-M. Chang, N. Tien, C.-T. Hsiao and C.-T. Peng (2011). "Building and Validating an Autoverification System in the Clinical Chemistry Laboratory." Lab Medicine **42**(11): 668-673.

Sikaris, K. (2012). "Application of the Stockholm Hierarchy to Defining the Quality of Reference Intervals and Clinical Decision Limits." The Clinical Biochemist Reviews **33**(4): 141.

Valdiguie, P. M., E. Rogari and H. Philippe (1992). "VALAB: Expert System for Validation of Biochemical Data." Clinical Chemistry **38**(1): 83-87.

Westgard, J. O., P. L. Barry, M. R. Hunt and T. Groth (1981). "A multi-rule Shewhart chart for quality control in clinical chemistry." Clin. Chem **27**(3): 493-501.