

Auszug aus dem Buch von Stavros Kromidas, Hans-Joachim Kuss (Hrsg.)  
„Chromatographische Messdaten richtig auswerten und bewerten“,  
WILEY-VCH, 2008

## **Über Bewertung und WERTung chromatographischer Daten**

Stavros Kromidas, Saarbrücken

### **Einleitung**

Als Dienstleister ist mir folgendes durchaus bewusst: Bei einer physischen oder virtuellen Zusammenkunft sollte der letzte Eindruck ein positiver sein – damit ist man der Zufriedenheit der Adressaten gewiss. So kritzelt man keinesfalls am letzten Tag eines Seminars um 16.15 Uhr noch schnell eine neue Differentialgleichung an die Tafel – mag sie auch noch so fundamental sein. Nein, man lässt das Ganze eher mit einem „Spielchen“ oder einer Gruppenarbeit ausklingen, die die Teilnehmer erfolgreich absolvieren können. So auch in einem Buch. Probleme, Schwierigkeiten usw. gehören in die vorderen Kapitel. Man schließt vernünftigerweise mit einem eher doch positiven Ausblick ab nach dem Motto „es gibt Probleme wie überall, aber irgendwie klappt es doch“ – das zustimmende Nicken der Leser ist dem Schreibenden recht sicher. Oder man fasst wenigstens das vorher Gesagte schlicht zusammen.

Ich werde diese „ungefährliche“, motivierende und letzten Endes oft sinnvolle Praxis hier nicht anwenden. Vielmehr werde ich nachfolgend versuchen, vereinfacht und wertfrei die Situation zu schildern, wie ich sie vor Ort in den Labors erlebe – auch auf die Gefahr hin, dass Unangenehmes zur Sprache kommt. Das ist weder Eigensinn noch ein Marketingtrick. Mir liegt auch nichts ferner als lediglich Kritik zu üben oder den Zeigefinger zu heben. Es geht vielmehr um eine – sicherlich subjektive – Darstellung der von mir erfahrbaren Alltagspraxis analytischer Labors, um daraus einige Rückschlüsse ziehen zu können.

### **Die Situation – oder warum ändert sich so wenig?**

Wenn im Privaten die Rede von „Bewertung“ eines Ergebnisses ist (ich habe seit gestern neue Fliesen in meinem Bad) geht es um eine kritische Stellungnahme zum „Wert“, d.h. um die Qualität des Ergebnisses, um jenes beurteilen zu können. Und ich berücksichtige bewusst oder erfasse zumindest intuitiv mehrere Aspekte des Ergebnisses (Preis, Farbe, saubere Fugen etc.). Aus dieser Erkenntnis heraus komme ich zu einem abschließenden Urteil, es könnte sich dabei evtl. auch ein Handlungsbedarf ergeben (z. B: Die Fliesen um die Badewanne herum sind noch einmal neu zu verlegen). Nun, wir wissen alle, dass im Berufsleben diese natürliche Haltung - in der Natur instinktiv vorhanden und damit im wörtlichen Sinne „natürlich“ - selten möglich ist bzw. geduldet wird. So reduziert sich in der Analytik die Bewertung oft auf eine „ja/nein“-Entscheidung eines Zahlenwertes. Es geht also nicht um den „Wert“ sondern lediglich um einen Zahlenwert. „Passt“ nun die Zahl oder „passt“ sie nicht?

Ferner wissen wir alle, dass es kein Problem ist, sie passend zu machen: Es kann jener Zahlenwert für die Totzeit in den Rechner eingegeben werden, um den in der PV geforderten k-Wert zu erhalten, Messungen werden wiederholt, damit der Wert hoffentlich wieder „in spec“ ist oder es kann manuell so integriert werden, dass die Peakfläche der Nebenkomponente gerade unter der geforderten 0,05% der Hauptkomponente liegt. Das ist zuweilen frustrierend oder hinterlässt - vor allem bei

den Anwendern direkt am Gerät - ein schales Bauchgefühl, das man irgendwann als unvermeidlich akzeptiert. Man ist auch unsicher, ob und was überhaupt „schlimm“ ist, auch die Strenge der Vorgaben ist nicht immer nachvollziehbar, z. B: An den Methodenparametern darf man kaum etwas ändern, aber bei der Integration ist man fast völlig frei, wie kritisch ist wirklich dieser kleine Peak, usw?

Warum ist die Situation so wie sie ist? D.h. warum beschäftigen wir uns so stark mit der Bewertung eines Zahlenwertes im Sinne der Übereinstimmung mit Vorgaben und so wenig mit der Wertung des Ergebnisses im Sinne einer analytischen Beurteilung? Stark vereinfacht können wir wie folgt festhalten:

In einer globalisierten, stark Wettbewerbs-orientierten Berufswelt fungieren Zahlen als einfache, nachvollziehbare und von allen Beteiligten akzeptierten Größen, ja, gar als Beweise für eine klar definierte „Qualität“. Wettbewerb fordert Vergleichbarkeit. Vergleichbarkeit verlangt nach objektiven Kriterien. Dazu eignen sich eben Zahlenwerte hervorragend. Je stärker nun der Wettbewerb ist, um so mehr gewichtet man Maßstabs-getreue Werte. Dadurch wird unvermeidlich einer wirklichen, analytischen Beurteilung entgegengewirkt. Woher die Zahlen stammen, ob sie repräsentativ sind, ob sie überhaupt eine Relevanz haben usw. ist oft sekundär und wird selten hinterfragt. Das ist ein großes Manko. In einer auf Vergleichbarkeit bedachte Analytik (Gesellschaft) herrscht schnell Zahlengläubigkeit. Deswegen steht weiter oben „Qualität“ und nicht Qualität, weil vor allem formale Aspekte zur Definition herangezogen werden. Die Wiederholbarkeit, das reproduzierbare Finden von Zahlenwerten wird als Beweis für gleichbleibende Qualität angesehen, es handelt sich jedoch bestenfalls um den Beweis von gleichbleibenden Eigenschaften – aber auch nur derjenigen, die überprüft werden.

#### *Bemerkung*

Jahrelang wurde im Labor dank GLP und anderen QM-Systemen auf Nachvollziehbarkeit und auf sture Befolgung von formalen Vorgaben getrimmt, deren Einhaltung gleich mit Qualität gesetzt. Man hat sich intensiv mit Qualitätssicherung statt mit Qualität befasst. Dass dies selbstverständlich langfristig nur ein Irrweg sein kann, hat die FDA längst erkannt und in den letzten Jahren einen vorsichtigen aber klar erkennbaren Kurswechsel eingeleitet: Statt einer „Any Change is Bad“- „Compliance, not Science“- und „Blind Compliance“-Haltung soll nun die Ära von „Science and Risk Based Compliance“ und „Continuous Improvement“ eingeläutet werden. Stichworte dazu: PAT, keine GMP-Checklisten sondern Qualitätssysteme, weniger Qualitätskontrolle mehr Qualitätssicherung.

Kommen wir zurück zu dem was ist und nicht was sein könnte oder was irgendwann vielleicht sein wird.

Im heutigen Umfeld – vor allem in stark reglementierten Bereichen - kann von Wertung gar keine Rede sein. Das ist auch gar nicht beabsichtigt und wird auch nicht verlangt. Also wird – obwohl wir alle wohl wissen, wie es geht – nicht unbedingt versucht, etwas Besseres, Sinnvolleres zu machen. Und im Leben, im Berufsleben insbesondere, machen wir selten etwas deswegen, weil wir es an für sich gut erachten. Wir sind nur dann bereit, etwas Festeingefahrenes zu ändern, wenn entweder unsere Sicherheit merklich erhöht oder das Belohnungszentrum im Gehirn aktiviert wird (Anerkennung, Freude, Lob etc. erfahren). Ansonsten lassen wir es lieber sein, schließlich wollen wir ja keine Revoluzer sein. Und da heute „Fehler machen“ gleich mit Schwäche gesetzt wird, ist die Angst vor Fehlern bis in die höheren Ränge von Politik, Wirtschaft und Gesellschaft sehr groß. Und wir leben bekanntlich nicht in einer Zeit, in der Helden hoch im Kurs sind. In einer Zeit fast permanenten Zeitdrucks, starker Betonung von Formalien, Beweisbarkeit- und

Alibimentalität usw., ist man froh, dass das Ganze einigermaßen „läuft“. Und es bleibt alles wie es ist.

### **Wie kann sich etwas ändern und wann ist es überhaupt notwendig?**

Am Anfang einer Entscheidung sollte „Können“ und „Wissen“ stehen. Erst wenn ich kann und weiß, erkenne ich Handlungsbedarf und kann Realisierungschancen und Wege beurteilen. Zum „Lernen“ und „Können“ sei hier nur folgendes angemerkt: Es sollte möglichst ein Klima geschaffen werden, in dem jene möglich sind. Von der Neurophysiologie wissen wir: Menschen lernen und arbeiten nicht unter Druck am besten, sondern, wenn sie sich wohl fühlen. In dieser (glücklichen) Situation wird der Neurotransmitter Dopamin aktiv; die Aufmerksamkeit steigt, das Gehirn ist bereit zu lernen. Ist diese Voraussetzung geschaffen, kann Wissen erlangt und gespeichert werden. Die Intention vorliegenden Buches ist es, das Wissen um die Auswertung in der Chromatographie und die Bewertung der erhaltenen Ergebnisse im Sinne einer analytischen Beurteilung ein wenig zu erweitern. Nachdem Wissen um den Wahrheitsgehalt von Zahlenwerten vorliegt und eine Wertung möglich ist, wird entschieden – vorausgesetzt die Ziele sind allen Beteiligten unmissverständlich klar - , ob Handlungsbedarf besteht und ob Handlungsspielraum vorhanden ist.

Dazu vier Beispiele:

- Der F-Test ist recht weit verbreitet; für einen tatsächlichen Vergleich der Streuung zweier Meßreihen ist er denkbar ungeeignet, da beim F-Test Varianzen (also Standardabweichung) und nicht relative Standardabweichungen - wo die Streuung in Bezug zum Mittelwert steht - verwendet werden. Wird er „expressis verbis“ vom Kunden oder von einer Behörde verlangt, ist er natürlich anzugeben. Es ist kein Zynismus sondern Realität, wenn wir festhalten: Das eine ist Analytik, das andere ist „Geschäft“.
- Ähnlich verbreitet und allgemein akzeptiert ist der Korrelationskoeffizient: Man muss sich nun sehr „anstrengen“, um im Falle von sauberen Proben keine zwei oder drei „9“ nach den Koma zu bekommen. Was bedeutet das? Mögliche Ausreißer und Matrixeffekte beeinflussen ihn kaum. Die relative Verfahrensstandardabweichung dagegen reagiert sehr empfindlich auf derartige Veränderungen. Möchte (muss ?) ich mir beim Routinebetrieb unabhängig vom Wissensstand der Anwender und sonstigen Variablen keine Probleme und Reklamationen einhandeln, bin ich froh, dass der Korrelationskoeffizient als Kriterium für die Linearität gefordert wird. Bzgl. diesem Punkt werde ich ein ruhiges Leben führen. Geht es tatsächlich um Wissen über die Qualität einer kalibrierbedürftigen Methode (Einfluß der Matrix auf die Empfindlichkeit, Streuung einzelner Werte) wäre die relative Verfahrensstandardabweichung zweifelsohne das bessere Kriterium.
- Eine HPLC-Methode, in der vier NebenkompONENTEN abgetrennt werden, wurde 1984 in mehreren Ländern eingereicht und registriert. In der Zwischenzeit kann man durch den Einsatz von selektiveren Säulen neun NebenkompONENTEN abtrennen, deren toxische Relevanz unbekannt ist. Die noch geltenden Anforderungen von 1984 werden weiterhin eingehalten, es besteht also kein akuter Handlungsbedarf. Es passiert auch tatsächlich nichts. Welches Wort ist hier angebracht: Bewertung, Beurteilung, Qualitäts(!)kontrolle?
- Führe ich die Validierung unter optimalen und nicht unter realen Bedingungen durch und berücksichtige dabei nicht alle Einflussfaktoren, erhalte ich natürlich kleine VK-Werte. Was ist nun meine Intention oder anders gefragt, an was wird die „Qualität“ meiner Arbeit gemessen? Erwartet mein Chef drei Validierungen in 6 Wochen, da Termine vorher vom Management festgelegt werden? Kann es sein, dass ich de facto gar keine Möglichkeit habe, das Testen gründlich(er) zu

machen? Muss (will) ich aus welchen Gründen auch immer mit kleinen VK's aufwarten? Dann ist die „Validierung“ wie oben beschrieben so durchzuführen. Die vorprogrammierten Probleme beim Methodentransfer, die unweigerlichen „OOS“-Situationen im späteren Routinebetrieb und die damit verbundenen recht hohen Kosten aus gesamtbetrieblicher Sicht usw. sind hinzunehmen und belasten schließlich die Qualitätskontrolle – nicht uns...

Die Beispiele lassen sich beliebig fortführen.

Eine bewährte Strategie zur Effektivitäts- und Effizienzsteigerung im Labor besteht aus drei Schritten. Zuerst kommt das völlig objektive, rationale, emotional nicht belegte Wissen. Dies führt zur Kenntnis von Fakten. Dann gilt es, Ziele zu definieren, zu erfragen bzw. zu vermitteln. Anschließend sollte man bei Bedarf die analytische Praxis unter Berücksichtigung der Möglichkeiten optimieren. In diesem dritten Teil der Kette sehe ich persönlich das größere Potential.

Auch hier einige Beispiele:

- Bei der Bewertung einer Methode sollte man daran denken, dass deren „Wert“ durch Kosten- und Zeitersparnis erhöht wird. Oft reichen simple, de jure erlaubte Maßnahmen, um eine Methode zu verbessern, z.B: Wenn schon nicht bei der Trennung, sollte wenigstens beim Spülen der Säule der Fluß von 1 ml/min auf 2 ml/min erhöht werden. Statt einer 4 mm könnte eine 3 mm-Säule verwendet werden. In beiden Fällen ergibt sich eine Ersparnis an Zeit bzw. an Eluentenverbrauch um ca. 50%.
- Man sollte den vorgegebenen Spielraum nutzen und dem heute allgemeinen Trend „aus Angst, preussischer als die Preussen sein“ entgegenzutreten. Die Behörden lassen es zu, ja sie fördern es teilweise sogar. Verkrustete Strukturen sind oft die eigenen. So sollten die erlaubten Modifikationen bei USP- und EP-Methoden tatsächlich genutzt werden (s. Kap. 3.3-3.5), im Falle von unbekanntem Einflussfaktoren auf mein Ergebnis sollte (muss) ich jenes mit dem Erweiterungsfaktor  $k$  multiplizieren und diese Zahl als die in der Routine erwartete/erlaubte angeben (s. Kap. 1.4).
- In der klassischen Statistik spielt der Faktor „Zeit“ keine Rolle. So wird im Alltag üblicherweise überprüft, ob sich der VK unter einem Schwellenwert und der Mittelwert innerhalb der vorgegebenen Grenze befinden. Also eine „ja/nein“-Entscheidung. Wird nun darüber hinaus der Mittelwert in Abhängigkeit von der Zeit mit Hilfe einer Kontrollkarte verfolgt, kann erkannt werden, ob eine zeitliche Abhängigkeit des Mittelwertes vorliegt oder nicht. Sollte dies der Fall sein, handelt es sich entweder um Alterungsprozesse bei der analytischen Methode (Chemikalien, Säule) oder um eine zeitliche Änderung des Gehaltes bedingt durch den Produktionsprozess (Vergiftung von Reaktanten, Zersetzungen, etc.). Mit Hilfe von Kontrollproben kann dies unterschieden werden. Sollte es an der Produktion liegen, ergibt sich womöglich ein großes Optimierungspotential, das durch ein einfaches analytisches Tool erkannt worden wäre.
- Man sollte sich in „worst case“-Mentalität und Hinterfragen üben, um zu realistischen Anforderungen zu gelangen. Wenn beispielsweise der VK meiner Methode 0,8% beträgt und die Spezifikationsanforderungen für den Gehalt mit 98-102% festgelegt wurden, ist völlig klar, dass einzelne Werte „OOS“ sein werden: Die Streuung der Methode ist zu groß im Vergleich zu den Spezifikationsanforderungen. In diesem Fall wäre – das wird übrigens von der FDA in den letzten Jahren immer mehr propagiert – die Anforderung 95-105% angebracht. Oder man nimmt in einem solchen Fall „Ausreißer“ nicht ernst, die Spezifikationsanforderung sollte nur für den Mittelwert gelten. Ferner gilt eindeutig zu belegen, um wieviel schlechter (belegbarer Qualitätsverlust und konkrete

monitäre Einbusse) das betreffende Produkt bei einer Forderung für die Streuung von z. B. VK=1,5% statt der aktuell geforderten 1% wird. Fehlt dieser Beleg, sollte man solche Anforderungen etwas großzügiger handhaben. Es gilt folgendem entgegen zu wirken: Durch unbedachte Anforderungen wird die Analytik teurer, ohne dass die Qualität des Produktes und damit sein Vermarktungswert um diesen Betrag zunimmt. Die immense Schwierigkeit allerdings, einmal formulierte Anforderungen durch inhaltlich richtige Argumente zu revidieren, ist mir geläufig.

### **Wer kann etwas ändern?**

Wird bei der Beurteilung einer Methode oder generell bei der Beurteilung der aktuellen analytischen Praxis ein Handlungsbedarf erkannt, wäre eine echte Änderung nur durch das obere Management möglich. Die Chance erachte ich jedoch als eher gering. Warum? Das Management ist weit weg von der Praxis, für dieses ist das Labor eine kostenintensive Notwendigkeit, die Zahlen zu liefern hat, um das Produkt/die Dienstleistung verkaufen zu können. Das Labor soll bitte schön solche Zahlen/Ergebnisse liefern, die gebraucht werden - nach dem Motto: „Sorgt dafür, dass es klappt“. Für das Management sind dies lediglich Zahlen, ähnlich wie andere Zahlen, mit denen es sich herumschlagen muss: Verkaufszahlen, Cash-Flow, Aktienkurse usw.

Ich zitiere einen Abteilungsleiter: „Ich sehe schon ein, dass die Höhenauswertung genauer ist, aber wir werden an unserer Praxis nichts ändern. Bringen Sie lieber meinen Leuten bei, wie sie vernünftig(!) über die Fläche auswerten können, damit das Ganze so bleibt, wie es immer war. Wenn es nicht anders geht, sollen sie die Analyse halt wiederholen. Das Einsetzen für die Höhenauswertung in allen Standorten würde mich sehr viel Zeit kosten – die habe ich nicht und schließlich habe ich bei Gott wichtigeres zu tun“. Keine befriedigende, aber eine durchaus nachvollziehbare Haltung.

Viele Anwender am Gerät wissen in aller Regel recht genau, was schief läuft - nur sie haben selten eine Entscheidungsbefugnis.

Veränderungen obliegen meist dem altgedienten Laborpersonal und dem Labormanagement: Diese Kollegen haben das fachliche Wissen und häufig auch mit etwas Mut, Kreativität und guten Argumenten die Möglichkeit, wenigstens kleine Veränderungen herbeizuführen. Dazu gehört das Überdenken von Spezifikationen bzw. von Kriterien abhängig davon, ob bei der aktuellen Methode formale Anforderungen oder eben analytische „Wahrheit“ im Vordergrund steht, s. Beispiele weiter oben. Lohnt es sich nun beim heutigen Arbeitspensum und Zeitdruck sich für etwas einzusetzen, wenn das Ganze mehr oder weniger doch läuft? Es liegt auf der Hand, dass die Antwort nur eine individuelle sein kann. Aus meiner Erfahrung würde ich vielleicht folgendes vorschlagen: Der erste, wichtige Schritt ist das Klarstellen von Fakten: Anzahl von Proben, vorhandene Ressourcen, warum braucht wer, bis wann, welches Ergebnis, mit welcher Präzision, usw. Daraus wird eindeutig ersichtlich, ob ein Handlungsbedarf besteht und wenn ja, zu welchem Grad dieser umsetzbar ist. Man sollte anschließend nur dann zur Tat schreiten, wenn damit persönliche Vorteile einher gehen: Höherer Bonus, Verstärkung der eigenen Position bzw. des Selbstbewußtseins, Anerkennung, persönliche Genugtuung, Schaffensfreude, Motivationsschub für die Gruppe usw. Ist es so, sind die Chancen auf Erfolg nicht einmal schlecht.