Rule 7. Watch out for the following frequent errors:

a. O for C and vice versa in the first member of a side chain.

b. Cyclopentyl or cyclohexyl rings with just a vertical double bond implying a methylene group where C=O is meant.

c. Careless usage of the solid, dotted, and wavy lines to show known or unknown stereo bonds. Once a paper uses a dotted or wavy line, one would expect that the solid lines mean something, but this is not always the case.

d. Beware of the 5-position in steroids. The hydrogen at this position can be either α(dotted) or β(solid). Authors often fail to indicate which it is, even where this can be determined from flow sheet or context.

e. Trade and nonproprietary trivial names often imply more than they say. Where an author says Demerol he means meperidine hydrochloride. Biochemical literature is loaded with articles referring to epinephrine where epinephrine bitartrate is meant. Recent issues of *CBAC* have made a false distinction between 2-isopropylamino-1-(2-naphthyl)-ethanol, Registry No. 54808, and its hydrochloride, No. 51025, under the false assumption that the trade name Nethalide refers to the base. It is the hydrochloride.

f. It is easy to confuse pyrrolidino with pyrrolidono.

g. In phenothiazines, the U. S. 3-substituent would often be called a 2-substituent in European literature. This confusing designation will be carried over to *CA* if the name appears in the title, but may be corrected in the *CA* Index. It will not be corrected in *CBAC*.

h. Italic lower case *p* is often confused with Greek rho, and rho with sigma.

i. In dosage statements, check milligrams *vs.* micrograms.

This list could be expanded almost indefinitely. At the risk of appearing to be overly cynical, I repeat Rule 1. Trust nothing.

If I have seemed unduly harsh in my criticism of current practice, I revert to humanistic literature and cite Wordsworth's "Ode to Duty (*11*).

"O, Duty. If that name thou love
. . . check the erring and reprove."

To which there should, of course, be an erratum. Dear Mr. Wordsworth:

"Love does not rhyme with reprove."

LITERATURE CITED

(1) Shakespeare, Wm., "Julius Caesar," Act 5, Scene 3, line 67.
(2) Pope, Alexander, "Essay on Criticism," line 525.
(3) Descriptive Price List, Winthrop Laboratories, Special Chemicals Dept., April, 1965.
(4) Private communication.
(5) Wagner, E. S., R. E. Davis, *J. Am. Chem. Soc.*, **88**, 7 (1966).
(6) McBee, E. T., W. L. Dilling, H. P. Brandlin, *J. Org. Chem.* **28**, 3595 (1963).
(7) *Ibid.*, p. 2255.
(8) *Ibid.*, **29**, 3744 (1964).
(9) Brace, N. O., *J. Org. Chem.*, **29**, 3744 (1964).
(10) *Ibid.*, **28**, 3102 (1964).
(11) Fieser, L. F., Fieser, M., "Topics in Organic Chemistry," Reinhold Publishing Corp., New York, 1950.
(12) Wordsworth, William, "Ode to Duty."

Procedures for Assessing Errors*

KENNETH L. COE, DON P. LEITER, Jr., HARRY L. MORGAN, and FERD R. WETSEL
The Chemical Abstracts Service, The Ohio State University, Columbus, Ohio   43210

User confidence in the validity of published chemical literature is directly related to the accuracy with which it is presented. The nature, types, and sources of errors which find their way into the primary and secondary journal literature are analyzed. A description of the various methods by which the effects of such errors can be minimized through abstracting and indexing techniques is presented.

The Chemical Abstracts Service (CAS) recognizes that many errors are introduced into primary and secondary journals in the publication processes. Rather than dwell on the present status, which will be reviewed briefly, we will discuss CAS plans for minimizing these errors through various abstracting, indexing, and computer-handling techniques.

At CAS, abstracts are prepared, edited, and indexed by subject and language experts. This processing locates and resolves many of the detectable errors or ambiguous statements which occur in the primary literature, and/or which occur in the abstracting and indexing operations. However, all such errors are not found even in these rigorous review and edit stages.

Since errors in any part of a paper or in the prepared abstracts—title, bibliographic reference, author names, body of the abstract, body of the paper, or even in the

indexes—can slip through to the published record, methods of correcting these errors have been established. For CAS, if an error in a published abstract is of major technical nature, the abstract is considered to be of no value, and it is reprinted correctly in *Chemical Abstracts* (CA) in its entirety; in such a case only the republished abstract is covered in the volume and/or the collective indexes. Thus, in retrospective searches the error-containing abstract cannot be located through the indexes.

Other errors are of a type which can be corrected by entries in our indexes. Errors in the title of a paper, in an author name, or in the journal citation are corrected in the printed author index. Errors in the body of the abstract may also be corrected in indexes. Figure 1 shows such a situation for the CA Subject Index.

Zirconium compounds. (See also *Zirconium alloys.*)
...toxy *tert*-butoxy chloro, formation

a...outlon and, 4264h
dicyclopentadienylzirconium bis(tetrahydroborate), **62:3638h** (Ref. should be to p. 1798.)
dicyclopentadienylzirconium chloride tetrahydroborate, **62:3639a** (Ref. should be to p. 1798.)
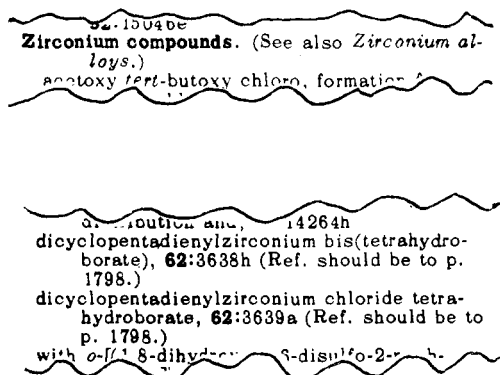with o-[..8-dihy...-disulfo-2-...

Figure 1. Chemical Abstracts Subject Index corrections to published abstracts.

The volume indexes are of substantial size, and under present methods of composition, typesetting, editing, and printing, the early parts of an index are completed before the latter parts of the index are completely processed. Errors in the early parts of an index detected while the latter parts are still in preparation are covered by errata listed at the end of that particular index. For errors detected still later, corrections are printed in the collective index for the period; errors detected after the publication of the collective index are corrected in the errata listed in the following collective index. Figure 2 is an example from a collective index. When CAS composition operations shift to a computer base, the editing of the entire index will be completed prior to composition; the more flexible computer composition procedures will permit corrections to be made to the index entries much later in the index generation process. In this way many of the errata in current volume and collective indexes will be eliminated.

CAS will within the next five years be a completely computer-based operation. At that time the final processing of *Chemical Titles* (CT), *Chemical-Biological Activities* (CBAC), and other contemplated new services will be fully integrated into the routine preparation of CA and its indexes. As CA is converted to a computer base, related operations will also be fully integrated so as to reduce multiple handling of the same material which can result in the introduction of errors.

One example of the integration of CAS procedures is the present movement toward single keyboarding of citations. All of the information for a citation—including title, author names and address, complete journal reference,

and the original language of the article—is input to the computer and stored there in one form. The information resulting from this single keyboarding can be manipulated by the computer so that the output can be in any form required for inclusion in any of a variety of publications. An additional example of this integration is the retrieval of indexed names for compounds previously recorded in the CAS Chemical Compound Registry System. The ability to retrieve names in this manner will eliminate renaming and editing names for compounds previously indexed, thereby reducing costs and the possibility of introducing errors.

The shift of production operations to the computer will allow many routine editing procedures to be carried out by computer programs. Two basic editing procedures are being programmed. In the first, the format is being analyzed so that it can be defined and the information is being analyzed for consistency and completeness. For purposes of illustration, consider the CAS Chemical Compound Registry System. When a structural formula is input to the computer it is accompanied by a molecular formula calculated by a chemist. The computer input is not verified, which saves much time; the computer is programmed to check the structural formula data against recorded standards. When the consistency of the elements of structural data has been established, the computer program recalculates the molecular formula and compares it with the one provided by the chemist. Should any discrepancy be detected, the compound is flagged for remedial action and re-input. Our experience in the processing of more than 300,000 structures indicates that the detection devices are over 99.7% effective on the total input for two-dimensional representation.

Programming the editing of structural formulas by computer is a great deal simpler than programming the editorial review of the average abstract. This is because the rules of preparation of structural diagrams are much more easily defined than the rules governing natural language. For the most part, we are at this time dependent upon humans to accomplish most editorial review of open text. Certain applications for review of spelling and hyphenation can now be assigned to the computer; however, the detection of inaccuracies and incompleteness in the text is generally left to the human.

An area which lies somewhere between the extreme difficulty of editing open text and the relative ease of editing structure data is the identification of inconsistencies in systematic chemical nomenclature and the translation of such nomenclature into more detailed structure descriptions. CAS is now working on computer programs to accomplish these purposes. By comparing the structural formula input to the Compound Registry with the record generated automatically from systematic chemical names, inconsistencies can be identified and brought out for review by a chemist.

The second procedure for detecting errors prior to computer storage is the building of machine checks into the data. An example is the Chemical Compound Registry Number which is a numerical address used for interlinking related data in various computer files. In copying numbers manually a simple interchange of two digits occurs frequently. Such errors are difficult to identify in a completely manual system, but can be avoided in the computer system

130

PROCEDURES FOR ASSESSING ERRORS

CORRECTIONS AND ADDITIONS*

SECOND DECENNIAL INDEX

(1917–1926)

SUBJECT INDEX

Page
2465 Under **Acetaldehyde**, from acetylene, **12**:1217[1], the vol. should be "**13**"; autoöxidation study of, in plants, **12**:741[3], the vol. should be "**13**."

2466 Under **Acetaldehyde**, collection of, P **16**:1251[2], the page should be "1252"; manuf. of, (*Patents*) **16**:1251[4], the page should be "1252."

2467 Under **Acetaldehyde**, water detection in, **15**:2601[5], the vol. should be "**14**."

Page
2486 Under **Acetic acid, chloro-**, manuf. of, P **16**:1251[5], the page should be "1252."

3906 Under **Ether**, α,β-**dichloroethyl ethyl**, **13**:2206[6], the page should be "2202."

5933 Under **Sodium perborate**, double salts contg., P **11**:8190[4], the page should be "1890."

THIRD DECENNIAL INDEX

(1927–1936)

SUBJECT INDEX

Page
3152 Under **Acetic acid, chloro-**, manuf. of, add "**25**:304[7]."

3877 Add "**1,2,3,4-Butanetetracarboxylic acid**, 1,4-dioxo-, tetraethyl ester, **27**:707[7]."

4808 Delete "**1,2,3,4-Ethanetetracarboxylic acid**, 1,4-diketo-' tetraethyl ester, **27**:707[7]."

4948 Under **Fireproofing**, compns. or agents for, (*Patents*) **21**: 4310[4,5], the vol. should be "**23**."

Page
5814 Under **Lupeol**, bombi-, from silkworm, **28**:2469[6], the column should be "3469."

6602 "**Phthaldehydic acid** (2-formylbenzoic acid)" should read "**Phthalaldehydic acid** (2-formylbenzoic acid)."

7912 Under **Wöhler, Friedrich**, urea synthesis by, 100th anniversary of, delete "**27**:5684[4]" and "**28**:3031[9]." Under **Wöhlerite** add "compn. of, etc., **27**:5684[4]; **28**:3031[9]."

FOURTH DECENNIAL INDEX

(1937–1946)

AUTHOR INDEX

Figure 2. Collective Index errata to previous indexes.

by adding a computer-generated digit. The Registry Number is a nine-digit number in which the ninth digit is a so-called "check digit" which results from programmed computation from the first eight digits of the assigned number. Thus the computer can detect most errors in the Registry Number. In addition to using the concept of a check digit in the Registry Number, a check digit has been incorporated in the CODEN reference. This CODEN reference is a compact, unambiguous abbreviation used to represent the title of serial, and some nonserial, publications. As a further check during the processing of CT and CBAC, the volume, issue, page, and year portion of the reference is checked by computer lookup to be certain that it is within the prescribed limits for the journal being processed.

Beginning with Volume 66 each CA abstract will be numbered separately. The abstract number will include a check digit. We hope that other publications will adopt a check-digit system for bibliographic citations to improve the reliability of machine-based systems for bibliographic retrieval. Patent numbers are a case in point in which check digits would be helpful. The check digit does not,

of course, eliminate the possibility of errors; however, it does minimize the probability of an undetected error. Since the check digit is an integral part of a code, it is not necessary that the reader know the rules by which the check is computed, though the existence of the check digit and the means of computing it should be made generally known to those who have computer systems. As a matter of fact, it is important that standards for check digits be established to increase the general utility of those which exist.

In practice, some level of undetected error must be assumed to be present in a store of information. To gauge the reliability and thus the utility of the store, it is desirable to make an effort to determine the extent and the nature of undetected errors. Such analysis is also obviously essential in order to provide for the improvement of the practices which led to the errors.

To determine the level of undetected errors in the CAS Chemical Compound Registry System on a statistically reliable basis, routine sampling procedures are used. In practice, the sample is completely reprocessed, independently of the initial processing, and the results of

the two processings are compared in detail. This procedure enables identification of the types, causes, and sources of undetected errors. The sampling procedure is not designed for the detection of errors. The fact that it does detect them is purely incidental.

As users of chemical information we understandably desire the elimination of all errors in both primary and secondary publications. However, as representatives of an organization which must recover the costs incurred in producing its publications, we must concede that some errors will exist in the published literature. CAS is striving, wherever possible, to reduce the number of errors in its publications and to make appropriate reference to errors which are detected in the primary literature we cover.

Control and Elimination of Errors in ISI Services*

IRVING H. SHER, EUGENE GARFIELD, and ARTHUR W. ELIAS
Institute for Scientific Information, 325 Chestnut St., Philadelphia, Pennsylvania 19106

The Institute for Scientific Information produces several indexes and abstracting and alerting services which contain error-controlling features. Curent Contents, Index Chemicus have different and unique mechanisms for finding and correcting errors that have appeared in the primary literature and those generated during their production of these secondary publications. Aspects of error control in these publications will be discussed.

Current Contents, Index Chemicus (IC), Science Citation Index (SCI), and ASCA (Automatic Subject Citation Alert) each have different and unique mechanisms for finding and correcting errors that have appeared in the primary literature and those generated during the production of these secondary publications.

It is obviously only possible to state very briefly some of the problems that exist, as well as some of the methods that are used to overcome them. This paper illustrates some types of errors involving people, other types involving machines, and, finally, general or composite errors.

The error-checking methods vary in complexity and extent according to the circumstances. For example, in the production of the Index Chemicus, not only are the articles indexed and graphically abstracted, but the molecular formulas of the individual compounds described by each author are recalculated by trained chemists. These abstracts are then sent to the original author for approval, giving him the opportunity of confirming corrected errors or making additional changes. Thus, the data found in Index Chemicus are sometimes more accurate than in the original article from which the abstract was prepared.

There is another important aspect to the whole problem of errors in the literature. Consider the perpetuation of a mistaken method or data that can go on being used for years without knowledge of subsequent modification. It is one of the unique capabilities of the Science Citation Index system that a straightforward check of the indexes will reveal such "corrections" (Figure 1).

An example of what one error in Current Contents can do is the case of an article by D. M. Baron of London,

1965 SCIENCE CITATION INDEX

D.G. Chapman, ANN. MATH. ST. 36, 1583 (1965).

CORRECTION TO
"A COMPARATIVE STUDY OF SEVERAL ONE-SIDED GOODNESS-OF-FIT TESTS"

By D..G. CHAPMAN

In the paper cited above (Ann. Math. Statist. 29 (1958) 655-674), it is stated that "any monotone test is admissible." This is in reference to the hypothesis $F = F_0$ against the alternative $F < F_0$. K. Doksum has pointed out that the test $\varphi = \alpha$ is a counter-example to this assertion which should therefore be deleted.

1583

Figure 1. Actual correction note by Chapman citing original 1958 paper.