データベースによる訂正記事抽出の問題点

向 田 厚 子*

金沢医科大学図書館

Mukaida A. (Kanazawa Medical University Library. 1-1, Digaku, Uchinada-machi, Kahoku-gun, Ishikawa-ken 920-0293, Japan) : Correction and Erratum in Databases. *Igaku Toshokan* 1999 ; 46 (1):62-69.

Correction (erratum) articles are usually found by screening journals, however, this is laborious and time consuming. Accordingly, a study to extract corrections by using databases (Current Contents and MEDLINE) was undertaken. The results showed that searching for them in the databases was not a perfect method, and features of correction information were clarified. Correction information is regulated by the latest manuals for authors and editors, although articles on corrections do not effectively function at present, only in electronic journals on the Internet. Finally, the future of correction information is anticipated.

Key words : Correction ; Erratum ; MEDLINE ; Current Contents ; Medical Writing Manuals ; Electronic Journals

I. はじめに

研究者は、研究成果を最終的に発表する場所として、学術雑誌に論文を投稿しています。出版物となるまでには、著者自身が何度も推敲し、編集者に査読され、完全な形となるはずですが、それでも出版されたあとになって間違いに気が付き、訂正が必要になる場合があります。それは一般的に後で「訂正記事」として、雑誌に掲載されます。

しかし、隅から隅まで雑誌を読み通し、注意をしていなければ、まずそのような訂正記事には気付きません。ほとんどの研究者は、論文をみつけることに精力をそそぎ、論文がのちに訂正されたかを確認し、把握する余裕などないと思います。

これまで図書館員は、訂正記事の重要性に着目

し、元の論文に訂正記事の情報を記載する作業を行ってきたり、訂正記事に関する様々な調査も行ってきました¹⁻⁶⁾。しかし、どの調査においても訂正記事を見つける方法は、雑誌の全ページに目を通すというスクリーニングで、多大な労力を要します。

最近では、文献検索に使う二次資料も、冊子体から電子化されたデータベースになり、瞬時に目的の文献を探すことができるようになっています。そこで、訂正記事を見つける作業をスクリーニングではなく、データベースからできないか試み、模索しました。ここでは外国雑誌を中心に、Current Contents (CC) と MEDLINE の二つのデータベースを使った場合の結果と、それらの調査を通して判明した問題点を、明らかにしたいと思います。

*Atsuko MUKAIDA : 〒920-0293 石川県河北郡内灘町大学 1-1.

atsuko-m@kanazawa-med.ac.jp

(1998 年 12 月 25 日 受理)

II. データベースの訂正記事データ

まず調査に入る前に、ここで使用する「訂正記事」の範囲を整理しておきます。孫野氏¹⁾は「正誤表」という表現を使用し、田井氏³⁾や山室氏⁴⁾は「訂正記事」という表現で、別刷りなどを含めた調査をしています。ここでは別刷りやシールなどを含めず、雑誌本体に掲載された訂正情報は、すべて「訂正記事」と表現して使用します。

それではまず、データベースにおいて訂正記事のデータが、どのような形で載っているかをみてみましょう。

1. Current Contents の訂正記事データ

CC の場合は、Document Type (文獻タイプ) が Correction となったものが訂正記事データを表していました (図1)。Source と Page は、訂正記事が掲載されている場所を表し、Title の括弧内は、元の論文が載っている場所を表しています。CC データベースはその名のとおりに、最新の雑誌における目次を毎週更新していますが、決して前のデータと結びつくことはありません。訂正記事データにおいても元の論文データと結びつきませんが、訂正記事データに元の論文の掲載場所が載っているので、それさえ把握すれば訂正記事の抽出作業に利用できると考えられます。

2. MEDLINE の訂正記事データ

MEDLINE の訂正記事データは、その訂正内容によって、違う表現を使っていることがわかりました。論文本体はそのまま生きているが、誤字・脱字など一部分だけを訂正する場合は、“published erratum” という表現になり、一般的な訂正記事に相当します。また、論文本体がなんらかの理由でもう一度出版された場合は、“corrected and republished” という表現で出ていることがわかりました。この場合、記事というより一つの

文献ですが、元の論文の訂正情報と考え、ここでは調査対象に含めます。そのほかに、元の論文を不正行為などの理由により撤回をする場合は、“retraction” や “retracted” という表現を使っています。そしてこれらは Title (TI:) フィールドでは大括弧で括った形で、Publication Type (PT:) フィールドでは大文字のハイフンで繋がれた形で載っています (図2-1~6)。

```

Author   GD Smith
Title    Mortality differences between black and white men in
         the USA: contribution of income and other risk
         factors among men screened for the MRFIT (vol 351, pg
         934, 1998)
Source   Lancet 351: 9112 (MAY 2 1998)
Page(s)  1366
Document Correction

```

図1. Current Contents の訂正記事データ例

```

TI:      HHV8 and skin cancers in immunosuppressed patients.
AU:      Boshoff-C; Talbot-S; Kennedy-M; O'Leary-J; Schulz-T; Chang-Y
SO:      Lancet, 1996 Jul 13; 348(9020): 138
PT:      PUBLISHED-ERRATUM

```

図2-1. MEDLINE の訂正記事データ例
(published erratum)

```

TI:      Carbon monoxide: a putative neural messenger [see comments] [published
erratum appears in Science 1994 Jan 7;263(5143):15]
CM:      Comment in: Science 1993 Jan 15;259(5093):309
AU:      Verma-A; Hirsch-DJ; Glatt-OE; Ronnett-GV; Snyder-SH
SO:      Science, 1993 Jan 15; 259(5093): 381-4
PT:      JOURNAL-ARTICLE

```

図2-2. MEDLINE の訂正記事情報が載った元の論文データ例 (published erratum)

```

TI:      Measurement error [corrected and republished in BMJ 1996 Sep
21;313(7059):744]
AU:      Bland-JM; Altman-DG
SO:      BMJ, 1996 Jun 29; 312(7047): 1654
PT:      JOURNAL-ARTICLE

```

図2-3. MEDLINE の訂正記事情報が載った元の論文データ例 (corrected and republished)

```

TI:      Measurement error [corrected and republished article originally printed
in BMJ 1996 Jun 29;312(7047):1654] [see comments]
CM:      Comment in: BMJ 1997 Jan 11;314(7074):147. Comment in: BMJ 1997 Jan
11;314(7074):147-8
AU:      Bland-JM; Altman-DG
SO:      BMJ, 1996 Sep 21; 313(7059): 744
PT:      CORRECTED-AND-REPUBLICISHED-ARTICLE; JOURNAL-ARTICLE

```

図2-4. MEDLINE の訂正記事データ例
(corrected and republished)

TI: Familial Alzheimer's disease associated with S182 codon 286 mutation [letter] [retracted by Chapman J, Korczyn AD, Goldfarb LG. In: Lancet 1996 Jul 20;348(9021):206]
 AU: Chapman-J, Asherov-A, Wang-N, Treves-TA, Korczyn-AD, Goldfarb-LG
 SO: Lancet. 1995 Oct 14; 346(8981): 1040
 PT: LETTER: RETRACTED-PUBLICATION

図 2-5. MEDLINE の訂正記事情報が載った元の論文データ例 (retracted)

TI: Retraction: familial Alzheimer's disease associated with S182 codon 286 mutation [retraction of Chapman J, Asherov A, Wang N, Treves TA, Korczyn AD, Goldfarb LG. In: Lancet 1995 Oct 14;346(8981):1040]
 AU: Chapman-J, Korczyn-AD, Goldfarb-LG
 SO: Lancet. 1996 Jul 20; 348(9021): 206
 PT: RETRACTION-OF-PUBLICATION

図 2-6. MEDLINE の訂正記事データ例 (retraction)

MEDLINE はこのような様々な表現のほかに、データ更新に特徴があります。CC と違い蓄積型だという特徴から、複雑な構造になっています。“published erratum” の場合、訂正記事データは最終的に元の論文に吸収されてなくなり（図 3-1）、その他の “corrected and republished” や “retraction” については、訂正記事データも元の論文データも残っていることが一般的ようです（図 3-2）。

また “published erratum” の場合、CC と違い訂正記事データには元の論文の記載場所が載っていません（図 2-1）。つまり、訂正記事データが入ってきた時点で抽出した場合、元の論文を結局は現物から確認しなければならず、役に立たないということが判明しました。元の論文データに訂正記事のデータが結びついた時点で、ようやく利用できるといえます（図 2-2）。

それに対し、“corrected and republished” や “retraction” は、元の論文データには訂正記事の掲載場所が、訂正記事データには元の論文の掲載場所が載っています（図 2-3～6）。つまり、それはどちらからも参照できるようになっていますが、そのようなデータの重複が、訂正記事の抽出作業を複雑にします。

図 3-1. MEDLINE のデータ更新例 1

図 3-2. MEDLINE のデータ更新例 2

Ⅲ. データベースによる訂正記事把握率

以上のことをふまえ、次にデータベースでどれだけ訂正記事を把握できるか調べました。

1. 所蔵タイトル対応状況

まず、それぞれのデータベースが、本学図書館で現在購読中の外国雑誌 563 タイトルに、どのように対応しているかをみました。

本学で使用している MEDLINE は、Silver Platter 社の CD-ROM 版で、CC はフロッピーディスク (FD) 版の Life Sciences (LS) と Clinical Medicine (CM) の二つのエディションです。調査のために使ったツールは、MEDLINE が “List of Journals Indexed in INDEX MEDICUS 1998” および “SERLINE” で、CC は “Current Contents Journal Coverage as of January 1998” です。

その結果、すべてのデータベースに対応している雑誌は、132 タイトルありました（図 4）。それぞれのデータベースの対応状況をパーセンテージ

図4. 図書館所蔵外国雑誌のデータベース対応図

図5. 図書館所蔵外国雑誌のデータベース対応率

ジで表すと、MEDLINE 87% (488 タイトル)、CC LS 66% (373 タイトル)、CC CM 36% (204 タイトル)、CC どちらか一方でも対応しているタイトルは、79% (443 タイトル) でした (図5)。

MEDLINE の対応誌となっていない13%に含まれていたものは、Review 誌や Year Book などのモノグラフシリーズと、医学以外の雑誌でした。また、CC は LS と CM の二つのエディションだと79%ですが、すべてのエディションでみると、対応率は82% (+16 タイトル=459 タイトル) になります。

2. 訂正記事の掲載度

次にそれぞれのデータベースにおいて、訂正記事データの掲載状況を調べてみました。

ここでは、①図書館所蔵外国雑誌 563 タイトル中、MEDLINE と CC の両方のエディションに対応している132 タイトルで、②総合医学雑誌の1997年のインパクト・ファクター4.0以上のもので、③刊行頻度が Weekly のもの、という条件を満たしている「BMJ」「JAMA」「Lancet」「N Engl J Med」の4タイトルで比較しました。1996年から1998年(1月～6月)までの3年間に各雑誌に掲載された訂正記事を対象にし、調査結果は1998年9月1日現在のものです。

図6は現物で確認した分を100%とし、それぞれのデータベースにおいて年ごとの掲載率を、グラフにしたものです。CCがどの年も90%前後な

図6. 年別掲載率

のに対し、MEDLINE は最近のものであればあるほど、掲載状況が悪いことがわかります。1996年分においても61%しかカバーしていません。NLMによると訂正記事が雑誌に載るまでのタイムラグは、平均4.7カ月だそうです⁷⁾。9月1日現在のデータということを考えると、MEDLINEで訂正記事に気付くのは、元の論文が出版されてから1年以上かかるということになります。3年間の平均をまとめると、CCは平均92%で、MEDLINEは平均38%になります。

3. まとめ

これらの結果をまとめると、MEDLINEはCCより本学図書館の外国雑誌をカバーしていることと、CCはMEDLINEよりも「訂正記事」のデータを載せていることがわかりました。つまり、データベースで訂正記事を抽出しスクリーニングの代わりに利用しても、完璧には把握できないことが判明しました。

IV. データベースの特徴

以上の調査から、それぞれのデータベースの特徴を、訂正記事の抽出作業の観点からまとめてみました。

1. Current Contents の特徴

CC FD 版の長所は、訂正記事データの掲載率が高いこと（平均92%）、データ更新が早いこと、データの形が一定していること、つまり Document Type が Correction のものだけ検索すれば良く、簡単であることが挙げられます。

短所は、LS と CM の二つのエディションだと、本学図書館の雑誌の網羅性が低いこと（79%）、FD 版は一括して調べられず、エディション別、号別でみなければならないことが挙げられます。Windows 版であれば、一括して 16 issue までみられる⁸⁾ので、Windows 版にすればこの短所は多少解消できると思われます。

2. MEDLINE の特徴

MEDLINE SilverPlatter 社の CD-ROM の長所は、本学図書館の雑誌の網羅率が高いこと（87%）、データを一括して調べられることが挙げられます。

短所は、訂正記事データの掲載率が低いこと（平均38%）、データ更新が遅いこと、データの形が一定でないこと、つまり erratum や correction や retraction で検索しなければならず、重複データにも気をつける必要がある、ということが挙げられます。

3. まとめ

訂正記事を抽出する作業にデータベースを用いるというアイディアは、調査を始めるまでは有効的な手段だと思えました。実際、完璧でないことを覚悟すれば、CC で訂正記事を抽出する作業は

楽になりそうです。しかし、その訂正記事データをダウンロードし、所蔵している雑誌の分だけを把握することは、決して容易なことではありませんでした。したがって、スクリーニングする代わりにデータベースを用いて訂正記事を抽出するアイディアは、図書館員の労力を省けないという結論に達しました。つまり、図書館で訂正情報を把握し、利用者に提供するというサービスは、現在でもなお困難な状況にあるということです。

しかし、データベースに訂正記事が掲載されているという事実は、少なくとも個々の図書館でスクリーニングする作業の必要性がなくなったといえます。田井氏³⁾によると、人間の目でスクリーニングした場合でも7割程度しか把握できないそうなので、CC のほうがずっと有効的です。訂正記事の掲載率を調査した時においても、データベースの情報から訂正記事に気付くことができました。

また、国内の医学関係データベース（医中誌 CD-ROM 版⁹⁾や JICST・医中誌医学文献ファイル）についても調べたところ、訂正記事データがないことがわかりました。このことから CC や MEDLINE は、質の高いデータを提供していることがわかります。

V. 訂正記事の多様性

そのほかに、先の調査結果を雑誌別にみた場合（図7）に、ある問題点に気付きました。他の雑誌

図7. 雑誌別掲載率

誌ではいずれも95%以上の掲載率を示していたCCで、N Engl J Med という雑誌だけ74%と、訂正記事の掲載率が異常に低いことがわかりました。その原因を探ると、訂正記事の多様性が浮かび上がりました。

N Engl J Med という雑誌の表紙には、目次が載っています。そしてそこに correction という項目があり、先の調査ではこれを訂正記事として件数に入れました。

目次の案内に従ってその訂正記事が載っている場所をみると、correspondence 欄であることがわかりました。correspondence 欄は、雑誌の編集者に宛てた手紙が載るところです。N Engl J Med という雑誌では、読者の指摘も訂正記事 (correction) として扱っていることがわかります。その記事についてデータベースでは、CC と MEDLINE の両方で、文献タイプを letter として扱っていました。

そのほかにも MEDLINE では、comment (CM:) というフィールドに、訂正情報が載っていることがありました。NLM では、comment フィールドを設け、元の論文やレターに対し、論争を挑んだり、論駁したり、支持したり、意見を追加したりする目的で書かれた論文やレターがあれば、そこに載せて関連付けています⁷⁾。

以上のことから、論文の誤りに気付く「訂正」する過程には、二つあることがわかります。一つは、著者や編集者が気付く誤りで、もう一つは、読者が気付く誤りです。そして、著者や編集者が気付く誤りは、一般的な訂正記事になり、読者が気付く誤りは、雑誌の correspondence 欄や letter 欄、commentary 欄、news 欄などに載るということです。そして後者の欄に載った訂正記事を把握することは、スクリーニングにおいても、データベースにおいても、とても困難であることがわかりました。

この部分は、訂正記事といえないのかもしれま

せん。しかし、これらの欄こそが科学のコミュニケーションにつながる、つまり科学の進歩につながる欄で、とても重要な部分に違いありません。N Engl J Med は、読者の指摘においても正当なものと評価した場合、correction として紹介していると思われます。

このように訂正情報の取り扱い方は、雑誌によって違い、またデータベースによっても違います。先の調査においても、データベースで訂正記事を完全に把握できない理由が、ここにあるのかもしれない。

VI. 訂正情報取り扱い規定

しかしながら、訂正情報の取り扱い方を統一しようという動きが、最新の医学論文の執筆マニュアルから窺えます。“Uniform requirements for manuscripts submitted to biomedical journals 5th ed” (生物医学雑誌への投稿に関する統一規定 第5版)⁹⁾ やアメリカ医師会の出した“AMA manual of style 9th ed”¹⁰⁾ では、著者が訂正記事を参考文献に挙げる場合の記載方法や、訂正情報に対する編集者の方針、加工の仕方などについて規定してありました。

MEDLINE では、1984 年頃から訂正情報の重要性を考え、データとして載せる方針をとってきました⁷⁾。“Uniform requirements...” においては4th ed (1991)¹¹⁾ より、参考文献で retraction, comment, erratum を挙げる方法が載るようになりました。“AMA manual of style” においても、8th ed (1989)¹²⁾ では MEDLINE の取り扱い方針が紹介され、9th ed でより詳細に規定しています。このように MEDLINE が方針を打ち立てたあと、そろって訂正情報関係の規定が載るようになったことから、二つの執筆マニュアルは、MEDLINE の影響を受けたことがわかります。

また、これらの執筆マニュアルから、訂正記事

の裏には、剽窃などによる不正行為や、重複出版と関係している場合があるとわかりました。特に論文の撤回については、科学研究における不正行為と深く結びついています^{13,14)}。先の年別掲載率の調査(図6)において、1998年のMEDLINEにおける訂正記事データは、すべて retraction であったことから、論文を撤回することが、いかに科学界に重要な影響を与えるかを物語っています。

VII. 訂正記事の機能達成度

このように、訂正記事には、ケアレスミスのようなちょっとした間違いから不正行為の発覚によるものまで、様々な内容のものがあることがわかりました。MEDLINEがCCと比べ訂正記事を網羅していないのも、内容の重要度を考慮し、取捨選択しているのかもしれません。しかしどの程度の訂正であっても、データベースにおいて訂正記事と元の論文を結びつけることは、重要だと思います。重要でないとすれば、それは雑誌においても同様に、掲載しなくても良いものといえます。つまり、訂正記事を掲載すべき程度のものかどうかについては、データベースの次元ではなく、雑誌の編集段階での問題だと思います。

以上の調査から、現在では完全でないにしろ、訂正情報がデータベースに集積され、執筆マニュアルでは、訂正情報の取り扱いを統一する試みがあるとわかりました。

では、実際にどれだけ科学界に、訂正記事の情報が浸透しているのでしょうか? 1990年から1997年にかけての生物医学分野における引用頻度の高いトップ10の著者¹⁵⁾が関わった論文で、後に訂正記事を出した分を、MEDLINEから抽出してみました。そしてそれらが Science Citation Index のデータベース SciSearch で、どれだけ引用されているか調べてみました。その結

果、一番多い文献は414件も引用されていましたが、その中で訂正記事を参考文献に挙げている人は、一人もいませんでした。訂正内容がただの誤字・脱字だけでなく、文章と参考文献の追加だったことを考えると、著者の論文を正確に把握していた人は、一人もいなかった恐れがあります。

この結果は、訂正情報が現在も、その機能を果たしていないことを表しています。訂正記事が掲載されたということは、少なくともその論文を参考にした研究者に伝わっていなければ、意味がありません。今後は雑誌の方針が統一され、データベースにすべて収録され、少しでも読者にその情報が伝わりやすくなることを願ってやみません。

VIII. 訂正記事の将来

しかし、このように訂正記事の役割が果たされていないという問題は、冊子体の雑誌が、インターネットにおけるオンラインジャーナルになった場合に、解決すると思います。オンラインジャーナルの一つである N Engl J Med では、訂正記事は元の論文とリンクされた状態になっていました。冊子体では役割が果たせていない訂正記事も、電子化された場合は、十分にその機能を発揮できることがわかりました。しかし、現在インターネット上で学術雑誌を読むことは、手軽な方法になっていません。冊子体の雑誌をベースにして電子化するという方法が、経済的な問題を引き起こしているからです。

Walker氏は、インターネット上で学術雑誌が無料でアクセスできるようになるモデル¹⁶⁾を紹介し、「もし完全電子化が行なわれれば、伝統的な雑誌は大きく変貌するであろう」¹⁶⁾と予言しています。それが現実となった場合には、訂正記事も同じように変貌することが予想されます。元の論文と訂正記事をリンクする方法から、HP更新のように、直接インターネットサーバに投稿し、訂

正版と古い版がそのままオンライン保管データとして残る¹⁶⁾という形になるのかもしれませんが。その予想は当たらないにしても、訂正記事の動向は、今後の学術論文の動向と深く結びついていくに違いありません。

本稿は、第11回北信越地区医学図書館員研修会(1998年10月5～6日 福井医科大学)および第5回医学図書館研究会(1998年11月11～12日 東京慈恵会医科大学)にて発表したものを基に、加筆修正しまとめたものである。

参考文献

- 1) 孫野しづ子. 正誤表について. 第9回医学図書館員研究集会論文集 1975 : 57-71.
- 2) 熊谷留理子. 洋雑誌におけるエラー情報の収集と問題点. 第8回医学図書館員セミナー論文集 1981 : 61-9.
- 3) 田井泰弘. 外国雑誌における訂正記事について. 第18回医学図書館員研究集会論文集 1983 : 79-88.
- 4) 山室春美. 和雑誌における訂正記事の現状. 第11回医学図書館員セミナー論文集 1984 : 74-89.
- 5) 隍美佐子. 外国雑誌の論文訂正について. 薬学図書館 1996 ; 41(1) : 15-8.
- 6) 中村雅子. 雑誌の訂正記事について. 第15回医学情報サービス研究大会抄録集 ; 1998年6月20-21日 ; 倉吉市. 医学情報サービス研究会 ; 1998.p.50.
- 7) Colaianne LA. Retraction, comment, and errata policies of the US National Library of Medicine. Lancet 1992 ; 340 : 536-7.
- 8) 棚橋佳子. 使いやすくなった Current Contents for Windows の特徴. 医学図書館 1996 ; 43(3) : 384-5.
- 9) International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals 5th ed. [引用 1998-09-20]. Available from : URL <http://www.acponline.org/journals/annals/01jan97/unifreqr.htm>
- 10) Iverson C et al. American Medical Association manual of style : a guide for authors and editors 9th ed. Baltimore : Williams & Wilkins ; 1997.
- 11) International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals 4th ed. BMJ 1991 ; 302 : 338-41.
- 12) Iverson C et al. American Medical Association manual of style. 8th ed. Baltimore : Williams & Wilkins ; 1989.
- 13) 山口直比古. 科学コミュニケーションと論文の Retraction. 医学図書館 1990 ; 37(2) : 93-9.
- 14) 山崎茂明. 生命科学論文投稿ガイド. 東京 : 中外医学社 ; 1996.
- 15) Superstars in Biomedicine, 1990-97. In : Holden C, editor. Most looked-to labs [news]. Science 1998 ; 280 : 1013.
- 16) Walker TJ (時実象一訳). 学術雑誌のインターネット上での無料アクセス提供. 情報管理 1998 ; 41(9) : 678-94.