

分担研究課題

小児慢性特定疾患データに関する疫学的検討

研究分担者 坂本なほ子 国立成育医療センター研究所 成育社会医学研究部成育疫学研究室長

研究要旨

小児慢性特定疾患登録データの精度管理のために、神経芽細胞腫と急性リンパ性白血病（ALL）の新規症例について病理登録データベースと照合し検討した。いずれについても、病理登録されている症例のうち約 50%が小慢登録されていた。残りの 50%が登録されなかった理由については不明であり、現在のところ、明らかにする手段はない。また、入力ミス等を明らかにする手段もなく、現時点では精度管理が実施できない状況である。小児慢性特定疾患登録データをデータベースとして研究利用するために精度管理は不可欠であり、今後、対策を講じていく必要がある。

研究協力者

羊 利敏

（国立成育医療センター研究所 成育社会医学研究部成育疫学研究室）

A. 研究目的

データは複数の情報源を利用することによって、把握漏れ、把握遅れ、誤情報を確認・補完する作業が可能となり、登録精度の評価も可能となる。この作業は、1レコードごとに複数の情報源からの情報を照合し、検討する必要がある。平成 20 年度は、小慢性特定疾患登録（以下、小慢登録）と神経芽細胞腫病理登録との間で、平成 21 年度は TCCSG 急性リンパ性白血病登録との間でデータを照合し、小慢登録率を検討した。

B. 研究方法

1. 資料

2005 年以後の小慢登録は生年月日などの情報が記載されているので、本研究では 2005 年から 2007 年の小慢登録データを使用した。

（ア）神経芽細胞腫

照合した神経芽細胞腫のデータベースは、調査対象とした医療機関の 1995 年から 2007 年における全ての神経芽細胞腫病理診断が登録されているものである。今回、そのデータベースから、検体受付日、性別、年齢（月齢）、原発巣、医療機関名のデータ提供を受けた。

（イ）急性リンパ性白血病

照合した ALL のデータベースは TCCSG (Tokyo Children's Cancer Study Group) の急性リンパ性白血病登録である。TCCSG 参加 50 施設のうち、本調査の内容・計画を理解した上でデータ利用許可が得られた 45 施設の匿名データの提供を受けた。TCCSG は急性リンパ性白血病に関して病理診断を行っており診断名の精度は 100%に近い。また、参加施設からの登録率も 100%に近い。今回はそのデータベースから、検体受付日、性別、年齢、医療機関名のデータ提供を受けた。

2. 方法

（ア）神経芽細胞腫

小慢登録から、登録年が 2005 年から 2007 年、診断名 = 神経芽細胞腫 or 神経芽腫、または ICD=C74.9、または ICD-O=9500 の新規申請で、

医療機関名が対象医療機関であるレコードを抽出した。以下の情報を用いて照合を行った。

- 性別
- 年齢・月齢±1か月
- 初診日≦検体受付日≦初診日+1か月
- 原発臓器名≡原発巣

照合に重視した順番も上記の順である。

(イ) 急性リンパ性白血病

2006年1月1日から12月31日に検体を受け付けた TCCSG データにおいて、下記の情報を用いて小慢登録から一致する症例を照合した。

- 診断名=急性リンパ性白血病
- 性別
- 年齢
- 初診日≦検体受付日≦初診日+1か月

照合に重視した順番も上記の順である。

なお、今回の調査対象を2006年のみに限定した理由は、小慢登録では年度区切りで登録されている可能性があるため中の年の2006年だけを対象とした。

C. 研究結果

(ア) 神経芽細胞腫

1. 小慢登録における一致レコード数

照合作業から表1の結果を得た。

表1 小慢登録における一致・不一致レコード数

小慢年度	小慢登録数	一致数	小慢登録での不一致数
2005	5	4	1
2006	3	2	1
2007	2	2	0

不一致であった2レコードについて詳細にデータを確認した。2005年に不一致であった1レコードは、初診年が平成10年(1998年)となっていた。また、2006年に不一致であった1レコードは初診年が平成13年(2001年)であった。

2. 病理登録における一致レコード

照合作業から表2の結果を得た。

表2 病理登録における一致・不一致レコード数

小慢年度	病理登録数	一致数	病理登録での不一致数
2005	9	4	5
2006	4	2	2
2007	5	2	3

各年に一致するレコードは約半数であり、全体として、18症例中8症例(44%)であった。

不一致であったレコードについてデータを詳細に確認した。2005年に不一致であった5レコードは1歳未満で性別不明の1レコード、1歳の3レコード、11歳の1レコードであった。2006年に不一致であった2レコードはそれぞれ1歳と2歳のレコードであった。2007年に不一致であった3レコードは1歳未満の1レコードと1歳の2レコードであった。

(イ) 急性リンパ性白血病

1. 病理登録における一致レコード数

2006年1月1日から12月31日に検体を受け付けた TCCSG データは160症例であり、その中で一致するレコードは78症例(48.8%)であった。

2. 小慢登録されている症例と未登録の症例

登録されている症例と未登録の症例の間の差異を検討した。登録されている78症例中、男性は40症例(51.3%)、女性は38(48.7%)であり、未登録で性別が明らかな75症例中、男性は45症例(60%)、女性は30(40%)であった。性差はみられなかった。

平均年齢に関しては、登録されている症例の平均年齢は6.1±4.5歳、未登録症例は5.9±4.1歳であり、平均年齢に差はみられなかった。年齢が明らかな152症例の年齢分布を表1に示す。

表 1 ALL 年齢分布

年齢	未登録(%)	登録済(%)	合計(%)
<1	8(47.1)	9(52.9)	17(100.0)
1	9(45.0)	11(55.0)	20(100.0)
2	12(52.2)	11(47.8)	23(100.0)
3	8(47.1)	9(52.9)	17(100.0)
4	5(45.5)	6(54.5)	11(100.0)
5	6(60.0)	4(40.0)	10(100.0)
6	2(40.0)	3(60.0)	5(100.0)
7	5(71.4)	2(28.6)	7(100.0)
8	1(25.0)	3(75.0)	4(100.0)
9	5(83.3)	1(16.7)	6(100.0)
10	4(50.0)	4(50.0)	8(100.0)
11	1(25.0)	3(75.0)	4(100.0)
12	2(25.0)	6(75.0)	8(100.0)
13	5(71.4)	2(28.6)	7(100.0)
14	1(25.0)	3(75.0)	4(100.0)
16	0(0.0)	1(100.0)	1(100.0)
合計	74(48.7)	78(51.3)	152(100.0)

D. 考察

情報源が単一であるデータベースは、シンプルで扱いやすい反面、その情報源の状況の影響を受けやすく、精度管理が難しい。複数の情報源を利用することによって、把握漏れ、把握遅れ、誤情報を確認・補完する作業が可能となり、登録精度の評価も可能となる。この作業は、1レコードごとに複数の情報源からの情報を照合し、検討する必要がある。いうまでもなく、一致に関しては照合のプログラムを組み自動化することが可能であり、不一致な場合に人間が個別に対応するしかない。

本小慢登録データベースは主治医が作成する医療意見書をもとに作成される。したがって、主治医からの報告が唯一の情報源となりがちで、登録漏れ、誤登録が生ずる可能性が高い。

神経芽細胞腫の検討において、小慢登録で不一致であったレコードはいずれも初診年が登録年

より5年前、7年前のものであった。これが、単なる入力ミスであるのか、それ以外の問題によるものであるのかは、今回の調査では不明である。しかしながら、生年を見ると、それぞれ初診年の前年、初診年と同じ年であり、生年と初診年の両方を誤入力した可能性は低いだろう。この2レコードの不一致が入力ミスであるならば、もっとも可能性が高い誤入力は、新規・継続申請の項目と考えられる。この新規・継続申請の項目は罹患把握には決定的な情報である。

病理登録で不一致であったレコードは、小慢登録されなかったレコードである。これも、今回の調査から原因を明らかにすることはできないが、年齢が1歳未満、性別不明といった属性から推測するに、登録前に死亡した可能性も高いと思われる。また、単に小慢医療費助成制度を利用しなかったことも考えられる。

通常、1年ごとの集計を行う際には、登録年ごとに集計する。今回の調査から、小慢登録データは、登録年に発生した登録件数の集計は可能であるが、登録年に発生した新規患者数の集計は不可能であることが明らかとなった。その理由は、新規・継続申請の誤入力の可能性が小さくはなく、初診年が登録年とは一致しないためである。また、病理登録されているのに小慢登録されなかったレコードが半分以上あったことも挙げられる。

ALLの結果では、ALL症例のうち半分が小慢登録され、半分が未登録であることが明らかとなった。未登録な理由を探索するため入手できた情報を比較した。

TCCSGからは診断名、性別、年齢、検体受付日、医療機関名という5つの情報しか提供されなかったため、性別と年齢以外の検討を行うことができなかった。性別、年齢に関して、小慢登録している症例と未登録の症例の差異は見られなかった。医療施設に関する偏りも、特には見られなかった。

小慢登録されない理由としては、登録について同意が得られない、乳幼児医療費助成制度を使用して小慢は利用しない、登録前に死亡してしまっ

たため等が考えられるが、今回のデータだけでは全く解明できない。

いずれの調査においても小慢登録率は 50%未満であり、この小慢データの利用範囲は限られているだろう。もし、このデータを利用するのであれば、未登録の理由を明確にしておく必要がある。

登録率が 50%であるデータから何らかの結論を導くには、疫学的観点から、残りの 50%の属性や登録されなかった理由について把握する必要がある。それは、登録された症例と登録されなかった症例の間に、結論に影響するバイアスが存在しないことを確認しなくてはならないからである。今後、この小慢データを公開するのであれば、様々な利用・解釈の上での制約を明記しなくてはならないだろう。

そもそも、小慢登録データは 1 レコードずつ、他のデータを照合させることはできず、集団としての集計にのみ使用すべきという意見もある。し

かしながら、1 レコードずつを検討できないのであれば、精度管理は不可能であり、データベースとしての価値は半減するだろう。

E. 結論

現在の小児慢性特性疾患登録データは、登録率も含め、精度管理がなされていない。今後、このデータを公開するのであれば、様々な利用・解釈の上での制約を明記し、周知しなくてはならないだろう。

F. 健康危険情報 なし

G. 研究発表 なし

H. 知的財産権の出願・登録 なし