

監査証跡データに基づくサンプリング SDV に関する研究

品質マネジメント研究

5215F004-5 大島一真
指導教員 棟近雅彦

A Study on the Method for Sampling SDV Based on Audit Trail Data

OSHIMA Kazuma

1. 研究背景と目的

製薬会社は、治験を医療機関に依頼している。医療機関では、医師や治験コーディネーター（Clinical Research Coordinator:以下、CRC）が、被験者から得られたデータを医療記録（以下、原資料）に残す。そして、原資料の中から、治験に必要なデータを症例報告書（Case Report Form:以下、CRF）に転記する。その際、治験の実施状況を確認する職種のモニターが、SDV（Source Data Verification）を実施する。なお、SDV とは、原資料と CRF の記載内容の一致を、目視で確認する作業を指す。

近年、医薬品の開発コストを低減させるため、SDV の効率化が唱えられている。従来は、CRF の全データに SDV を実施しており、膨大な時間を費やしている。そこで、各データの重要度を考慮し、SDV を行う項目や頻度を絞るというサンプリング SDV が検討されている。しかし、標準的な方法は確立しておらず、経験的に SDV を行う項目や頻度を絞り、サンプリング SDV を実施している。

そこで本研究では、電子 CRF へのデータ処理が時系列に記録されたデータ（以下、監査証跡データ）を用いて、誤りの発生頻度や影響度を分析し、その結果をもとに、サンプリング SDV の設計方法を提案することを目的とする。

2. 従来研究と本研究のアプローチ

2.1. 従来研究

日本製薬工業協会の臨床評価部会[1][2]は、SDV の効率化に向けた取り組みとして、重要度に基づきデータを選択するサンプリング SDV を提案している。この方法では、まず、対象の試験におけるデータの重要度を評価する。そして、重要度が高いデータは全症例 SDV を行い、重要度が低いデータは SDV の頻度を減らすという方針をとる。また、重要度が高いデータとして、「患者の存在」、「同意取得」、「有害事象情報」、「主要評価項目」、「適格性」の 5 項目が挙げられている。

しかし、従来研究では、重要度が高いデータをどのように選択したかのプロセスが明らかではない。そのため、SDV が必要とされるデータ（以下、SDV 対象データ）の漏れや、SDV の不必要なデータが、過度に選択されている可能性がある。

2.2. 本研究のアプローチ

本研究では、まず、サンプリング SDV の設計方法の検討にあたり、CRF に入力されたデータの誤り（以下、データの誤り）に着目する。なお、データの誤りの例として、原資料から CRF への転記ミスや、原資料と CRF の記載内容が異なることなどが挙げられる。そして、製造業にお

ける検査設計の考え方を参考に、SDV 対象データを選択する際に、「データの誤りの発生頻度」と「データの誤りが、治験の評価や信頼性に与える影響度（以下、データの影響度）」の 2 項目を考慮する。

つぎに、これらの事項を考慮して、SDV 対象データを選択する方法を検討する。この際、「データの誤りの発生頻度」を明らかにするため、過去の治験で得られた監査証跡データを用いて、データの誤りを分析する。そして、この分析結果より、サンプリングの単位を決定する。一方、「データの影響度」を明らかにするため、製薬会社の治験の専門家への調査を行う。以上をふまえ、サンプリング SDV の設計方法を提案する。

3. サンプリング SDV の設計方法の提案

3.1. SDV 対象データ選択時に考慮すべき事項の検討

サンプリング SDV の設計は、どの項目を、どのような方法でチェックし、どう判定するかを決める必要がある。これは、製造業における検査の設計と同様といえる。そこで、検査設計に関する文献調査の結果から、SDV 対象データを選択する際に考慮すべき事項を検討した。

製造業の検査設計では、検査対象を決定する際、「不良の発生頻度」、「不良が製品に与える影響度」、「不良の検出難易度」の 3 点を明らかにすることが必要とされている。

そこで、検査で見つかる製品の不良を、SDV ではデータの誤りと捉えた。そして、SDV 対象データの選択時に、「データの誤りの発生頻度」と「データの影響度」の 2 項目を考慮することにした。この際、「データの誤りの検出難易度」は SDV では評価は変化しないと考え、本研究では除外した。

3.2. 監査証跡データの分析方法の検討

3.1 節で示した SDV 対象データを選択する際に考慮すべき事項のうち、「データの誤りの発生頻度」を明らかにするため、監査証跡データを分析した。なお、分析に用いた監査証跡データの概要を表 1 に、監査証跡データの一例を表 2 に示す。

表 1. 用いた監査証跡データの概要

	薬効	用いた 症例数	Excelの行数
試験 α	がん	計18症例 (全2施設分)	246286行
試験 β	高血圧	計6症例 (3施設×2症例)	10094行
試験 γ	がん	計17症例 (全5施設分)	278599行
試験 δ	インフルエンザ	計20症例 (10施設×2症例)	59890行

表 2. 監査証跡データ

施設	Form名	Field名	ログ番号	アクション	誰が	時刻
A	バイタルサイン	体温	0	「36.7」を入力した.	CRC	2015/6/11 5:47
A	バイタルサイン	脈拍数	0	「96」を入力した.	CRC	2015/6/11 5:47
A	手術歴	開始日	0	SDVをした.	モニター	2015/6/11 5:51
...	...	...	...	...	...	...

監査証跡データは「施設名」や「Form 名」など 13 列で構成されており、CRF へのアクションがとられた時刻ごとに記録される。なお、CRF の Form とは、CRF のページ名であり、数項目の Field から構成される。

この監査証跡データは、異なる Form や Field に対するアクションが、単に時系列順に記録されるだけである。そのため、特定の入力されたデータが固定（ロック）されるまでの経緯を辿ることは困難であり、このままの形ではデータの誤りの分析が行えない。そこで、入力されたデータの一連のプロセスを切り出せるように並び替えた。

表 3. 切り出したデータ群（一部）

施設	Form名	Field名	ログ番号	アクション	誰が	時刻
A	バイタルサイン	体温	0	「36.7」を入力した.	CRC	2015/6/11 5:47
				「36.9」を入力した.	CRC	2015/8/26 9:43
				データをフリーズした.	モニター	2015/8/28 1:22
				データをSDVした.	モニター	2015/9/3 1:37
				レビュー内容に問題がないことを確認した.	DM	2016/1/10 8:38
				電子署名をした.	Pir	2016/2/15 1:09
				データをロックした.	DM	2016/2/24 4:23

表 3 のように、監査証跡データの「施設名」から「ログ番号」までの列が同じ内容のとき、入力されたデータの一連のプロセスを切り出せたことになる。

つぎに、切り出したデータ群の中から、データの誤りを抽出する。そこで、切り出したデータ群の中から、データの誤りの可能性がある変更があったデータを抽出し、それらを以下の 3 通りに分類した。

(ア) クエリ発行で見つかったデータの誤り

データに疑義がある場合に、CRC などへの問い合わせとしてクエリが発行される。(ア)は、このクエリによって見つかったデータの誤りである。なお、クエリは事前に設定されたプログラムによるチェックや、モニターやデータマネージャーによるマニュアルを用いたチェックなどによって発行される。監査証跡データ上では、クエリと分かるフラグが示されている。

(イ) CRC が自ら気付いたデータの誤り

これは、CRC の見直しによって、自ら気付いたデータの誤りである。本研究では、監査証跡データ上で、入力されてから 1 日以内に更新されたデータと仮定した。

(ウ) SDV によって見つかったデータの誤り

これは、モニターが実施した SDV によって見つかったデータの誤りである。本研究では、監査証跡データ上で、入力されてから 2 日以降に変更されたデータと仮定した。また、その中でも、空欄だった項目に新しくデータが入った場合と、治験中にデータが新しく更新されたと明らかに考えられる場合は除外した。

これらの 3 つの分類の中で、(ウ)がサンプリング SDV で見つけるべき誤りと考えられる。そこで、切り出した一連のプロセスの中に、(ウ)があるかを判断する作業を繰り返し、この発生件数を集計した。

3.3. サンプリングの単位の検討

3.2 節の監査証跡データの分析方法を用い、データの誤

りの発生頻度を求め、その結果から、サンプリングの単位を決定する。まず、データの誤りの発生頻度を求めるため、表 1 の分析対象について、試験、施設、症例、CRF の Form、Field の単位でデータの誤りの件数を集計した。なお、データの誤りの発生割合は、データの誤りの件数を分子、入力されたデータの総数を分母として算出した。結果の一部を表 4 に示す。

表 4. データの誤りの発生頻度（一部）

試験	施設名	症例名	データ数	症例単位		施設単位		試験単位	
				誤りの 件数	発生割合	誤りの 平均件数	発生割合	誤りの 平均件数	発生割合
α	V	V-1	4806	201	4.2%	121.4	4.58%	123.9	5.56%
		...	...	...	...				
		W-1	1825	200	11.0%				
β	X	X-1	549	1	0.2%	1.0	0.29%	1.5	0.48%
		X-2	152	1	0.7%				
		...	...	...	...				
γ	A	A-1	4577	145	3.2%	101.8	3.53%	72.6	2.43%
		A-2	2855	134	4.7%				
		...	...	...	...				
δ	J	J-1	646	0	0%	0	0%	1.5	0.22%
		J-2	646	0	0%				
		...	...	...	...				

表 4 より、試験、施設、症例単位ごとのデータの誤りの発生件数と発生割合が明らかになった。試験単位で見ると、薬効が「がん」の試験 α と γ のデータの誤りは比較的多く、試験 β と δ はほぼ発生していなかった。また、施設、症例単位では、それぞれ多少の値の違いはあるが、データの誤りの出方に傾向は見られなかった。

つぎに、試験 γ の Form 単位でデータの誤りを集計した結果を示す。

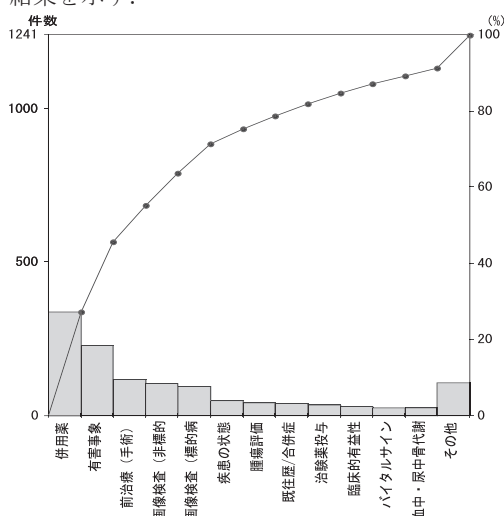
図 1. 試験 γ の Form 単位集計時のパレート図

図 1 より、Form 単位で集計した場合、試験 γ の全 37Form 中 8Form で累積 80%を超えていた。また、試験 α も同様の結果が得られた。一方、CRF の Field 単位で集計した場合は、傾向が見られなかった。以上より、サンプリングは、集計時に傾向が見られた Form 単位とした。

3.4. データの影響度の検討

3.1 節で示した考慮すべき事項のうち、「データの影響度」を検討した。この際、試験 α を事例とし、製薬会社の治験の専門家に対し、データの影響度を調査した。なお、3.3 節の結果より、データの影響度の評価は Form 単位とした。また、必ず SDV すべきデータを 5 点、SDV をほぼ

必要としないデータを1点の5段階評価とした。さらに、各評点を付けた理由も調査した。結果の一部を表5に示す。

表5. 試験αのデータの影響度の評価結果(一部)

影響度	Form名	評価理由	Form数
5	System Enrollment	症例の存在を確認するため	10
	Adverse Events	本試験の主目的のため	
	...	...	
4	Target Tumor Assessments	探索的評価項目のため	8
	...	...	
	...	...	
3	Demographics	適格性関連項目のため	16
	Prior Cancer Therapy		
	...		
2	Post Study Treatment	解析に使用されないため	7
	Prior and Concomitant Medications	重要ではないため	
	...	...	
1	Continuation	データの操作制御用のため	4
	...		

表5より、試験αの各Formの影響度が明らかになった。さらに、別の試験の評点をつける際の観点を明確にするため、評点ごとの評価理由を整理し、表6の評価基準を作成した。

表6. 評価基準(一部)

評点	評価基準
5	有害事象のデータである
	適格性の判定結果のデータである
	主要評価項目である
	...
3.4	探索的評価項目である
	有害事象を除く安全性評価項目である
	...
	...
1.2	解析に必要なないデータである
	EDCの操作制御用のデータである
	...

表6に示すように、影響度は5段階で評価されたが、今回は1試験分の評価結果であり、評点3、4と評点1、2の違いに明確な基準を設けることは困難だったため、実際の評価は3段階で行うことにした。今後は、複数の試験の評価結果を蓄積することで、より精緻な評価基準が作成できると考えられる。

3.5. サンプリング SDV の設計方法

STEP1: 提案方法を適用する試験の決定

サンプリング SDV を適用する試験を決定する。また、この適用対象の試験と薬効が類似する、過去の試験を選択する。

STEP2: CRF の共通 Form の把握

サンプリング SDV を適用する試験と、過去に実施した試験で共通する Form を把握する。Form 名が一致していない場合でも、Form の記載内容が試験間で対応していれば共通しているとする。

STEP3: データの誤りの発生件数の集計

過去に実施した試験の監査証跡データから、3.2 節の分析方法と定義を参考に、SDV によって見つかったデータの誤りの発生件数を Form ごとに集計する。

STEP4: データの影響度の評価

3.4 節の評価基準を参考に、Form ごとにデータの影響度を5段階で評価する。

STEP5: SDV 対象データの選択

データの誤りの発生頻度に関わらず、影響度5点のFormはSDV対象、影響度1,2点はSDV対象外とする。影響度3,4点かつ、STEP2で共通としたFormは、過去の試験で誤りの件数が多かったFormからSDV対象データとする。なお、累積何%までのFormをSDV対象とするかは試験ごとに設定する。また、STEP2で共通しなかったFormもSDV対象とする。

4. 事例への適用と効果の検証

4.1. 事例への適用

STEP1. 提案方法を適用する試験の決定

試験αは、試験γと薬効が同じであり、誤りの発生頻度も同程度であることから、試験γを過去に行った試験として、試験αに対して提案方法を適用することにした。

STEP2. CRF の共通 Form の把握

試験αと試験γのCRFにおいて、共通するFormを把握した。この際、Form名が一致している場合以外にも、Formの記載内容ごとに大分類を設け、共通しているかを判断した。その結果を以下の表7に示す。

表7. 試験αとγのForm対応表(一部)

分類	Form名	
	試験α	試験γ
システム登録	System Enrollment	プライマリーフォーム
被験者背景, 同意取得	Informed Consent	被験者背景
	Demographics	
バイタルサイン	Vital Signs	バイタルサイン
適格性	Subject Eligibility	適格性基準
	Subject Enrollment	
...	...	...
有害事象	Adverse Events	有害事象
再同意	Re-informed Consent	再同意
総合コメント	General Comments	総合コメント

表7より、試験γのFormが、試験αの全45Form中25Formと対応付いた。

STEP3. データの誤りの発生件数の集計

試験γの監査証跡データから、3.2 節の分析方法と定義に基づいて、SDV によって見つかったデータの誤りの発生件数を表7の大分類ごとに集計した。

STEP4. データの影響度の評価

3.4 節で、試験αのFormごとにデータの影響度を評価したため、その結果を用いた。

STEP5. SDV 対象データの選択

STEP4で影響度5点となった10FormをSDV対象データ、影響度1,2点となった11FormをSDV対象外とした。そして、影響度3,4点かつ、STEP2で共通としたFormにおいて発生した、SDVによって見つかったデータの誤りのうち、累積90%以内をSDV対象データとした。その結果を以下の図2に示す。

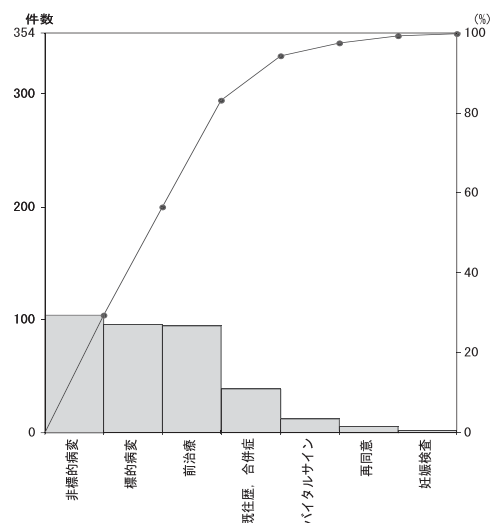


図2. データの誤りの件数のパレート図

図2の「非標的病変」から「既往歴，合併症」の大分類までの6FormがSDV対象データとなった。さらに，STEP2で共通とされなかった13Formを加え，試験αでは計30FormがSDV対象データとなった。

4.2. 効果の検証

提案法の有用性を確認するため，比較対象として既存のサンプリング案（以下，既存案）を挙げ，4.1節と同様に試験αに適用した。既存案の概要は以下の通りである。

- ・ 治験の実施設をA，B，Cの3段階で評価する。なお，Aが良い施設，Cが悪い施設である。そして，評価結果より，各施設のSDVの方法を変更する。
- ＜A評価の施設＞
症例の1/4は，影響度3,4,5点のFormをSDV。
症例の3/4は，影響度5点のFormをSDV。
- ＜B評価の施設＞
症例全て，影響度3,4,5点のFormをSDV。
- ＜C評価の施設＞
症例全て，全項目SDV。

なお，A評価の施設の場合の「症例の1/4」は，当施設における症例をランダムに5回選択し，それぞれの結果の平均値を用いた。

提案法と既存案の結果を比較する際，良し悪しを比較するための評価指標が必要である。そこで，3.1節同様，製造業の検査を参考に，検査の良し悪しを評価する際に考慮される労力と検出力に着目し，以下の表8に示す評価指標を5つ挙げた。

表8. 評価指標

評価指標
(1) SDV対象データとなったForm数
(2) SDV対象データとなった，データの影響度1, 2点のForm数
(3) SDV対象データとなった，データの影響度5点のForm数
(4) 検出できたデータの誤りの件数
(5) データの影響度で重みづけをした，検出できたデータの誤りの件数

表8の(1)と(2)はサンプリング案の労力を測る指標で，値が小さい方が良い。(3)(4)(5)は検出力を測る評価指標である。なお，(5)は影響度の高いデータの誤りが多く検出できたかを評価する指標である。

これらの指標を，提案法と既存案の事例適用の結果から求めた。また，全項目SDVした場合を100%とし，それに対する割合を算出し，表のカッコ内に示した。表9に結果の一部を示す。なお，既存案は，試験αの施設VとWの2施設がA，B，C評価だった場合の全9通りを適用しており，例えば[AC]は施設VがA評価，施設WがC評価の場合を表している。

表9. 提案法と既存案の適用結果（一部）

評価指標	本研究	[AA]	[AC]	[CC]
(1)	540(67%)	300(37%)	613(76%)	810(100%)
(2)	0(0%)	0(0%)	121(61%)	198(100%)
(3)	180(100%)	180(100%)	180(100%)	180(100%)
(4)	1829(82%)	1036(46%)	1790(80%)	2231(100%)
(5)	485(79%)	340.2(55%)	522.8(85%)	615(100%)

表9の[AA]が最も積極的にサンプリングを行った場合，[CC]はサンプリングを行わず，全項目SDVをした場合である。まず，[AA]と比較すると，労力の面は劣るが，検出力は明らかに大きな値である。さらに，検出力の面で提案法と近い値となった[AC]と比較した場合，提案法の方

が労力の面で良い値であり，効率的にデータの誤りが検出できている。以上より，提案法は，労力と検出力の両面を考慮できた提案といえる。しかし，今回は1事例での検証であるため，複数の試験に適用した結果を確認する必要がある。

5. 考察

従来では，データの重要度の捉え方が明らかではなく，経験的に重要と考えられるデータをSDVの対象としていた。そこで本研究では，SDV対象データを選ぶ際に考慮すべき事項を，検査設計の考え方を利用し，明らかにした。そして，考慮すべき事項である「データの誤りの発生頻度」と「データの影響度」を基準とすることで，SDV対象データを一意に選択することを可能とした。

監査証跡データは，CRFの作成の経緯をすべて記録したものであり，データの信頼性を担保するために用いられる。そのため，監査目的以外で，このデータの分析はされていなかった。そのため，どのような視点や単位で分析すれば，サンプリングSDVの設計に活かせるのかは明確でなかった。そこで，監査証跡データから抽出可能な「データの誤り」に着目した。そして，データの誤りを様々な単位で分析することで，発生頻度に違いが表れる単位を明らかにし，過去の知見を活用したサンプリングSDVを実施できるようになった。これより，監査証跡データの新たな活用方法を明らかにした。

本研究では，監査証跡データの分析方法を用いて，考え得るサンプリングSDVの候補の中から，データの誤りの出方に傾向がある，Form単位のサンプリングに決定した。施設単位や症例単位のサンプリングでは，データの誤りの出方に傾向がないため，データの誤りの発生頻度は施設や症例に依存してしまう。これに対し，提案法のForm単位のサンプリングでは，誤りの発生頻度が多いデータから，SDV対象を選択できる。さらに，データの影響度も考慮することで，重要度が高いデータを効率よく選択することが可能である。以上の理由より，提案法の有用性が高くなったと考えられる。しかし，この結果は，4試験分の監査証跡データの分析に基づいたものであり，今後分析対象を広げ，Form単位のサンプリングの妥当性を確認する必要がある。

6. 結論と今後の課題

本研究では，SDV対象データを選択する際に考慮すべき事項を挙げ，監査証跡データの分析や製薬会社への調査から，それらについて明らかにした。また，その結果に基づいたサンプリングSDVの設計方法を提案した。そして，実際に試験に提案法を適用することで，その有用性を確認した。

今後の課題は，分析対象とする試験数を増やした際の，サンプリングの方針の再検討が挙げられる。

参考文献

- [1] 日本製薬工業協会医薬品評価委員会臨床評価部会(2009)：“SDVの効率化検討”
- [2] 日本製薬工業協会医薬品評価委員会臨床評価部会(2013)：“モニタリングの効率化に関する提言”