

監査証跡データを用いるサンプリング SDV の設計方法に関する研究

品質マネジメント研究

5218F014-1

坂尻雄飛

指導教員

棟近雅彦

A Study on a Method to Design Sampling SDV Using Audit Trail Data

SAKAJIRI Yuhi

1. 研究背景と目的

治験とは、厚生労働省から新薬製造販売承認を受けるための、実証データの収集を目的とした臨床試験（以下、治験）であり、製薬企業は治験を医療機関で実施する。医療機関において、被験者から得られるデータは、患者カルテ等の治験の事実経過の再現と評価に必要な記録である原資料に残された後、症例報告書（Case Report Form：以下、CRF）に転記される。その際、治験の実施状況を確認するモニターが、治験データの信頼性確保のための業務の一環として、原資料と CRF の記載内容の一致を目視で確認する作業（Source Data Verification：以下、SDV）を実施する。

従来は、CRF に入力された全てのデータに対して SDV が実施されており、膨大な労力を要していた。そこで、治験プロセスの効率化のため、CRF に入力されたデータをサンプリングして SDV を行う、サンプリング SDV（以下、S-SDV）が検討されている。しかし、各製薬企業が試行錯誤的に S-SDV を設計・導入しているのが現状であり、標準的な S-SDV のアプローチは存在しない。そのため、データの信頼性保証に対する品質管理の考え方に基づいた、S-SDV の設計方法を確立する必要がある。

本研究では、電子 CRF へのデータ入力処理が、正確なタイムスタンプとともにシステムが自動的に記録する監査証跡のデータ（以下、監査証跡データ）を分析し、CRF 入力時に誤って入力されたデータが訂正される頻度（以下、訂正頻度）を検討することで、S-SDV を設計する。そして、その際に考慮すべき事項を手順化することにより、S-SDV の設計方法を確立することを目的とする。

2. 従来研究と研究方法

2.1. 従来研究とその問題点

日本製薬工業協会 [1] は、SDV の効率化に向けた取り組みとして、データの重要度を考慮した S-SDV を提案している。この方法では、対象治験のデータのうち、重要なデータに対しては全症例で、その他のデータについては症例数を絞って SDV を行う。しかし、データの重要度の評価方法等の、具体的な S-SDV の実施方法を明らかにしていない。

Rutger [2] は、データの誤りの発生頻度を考慮し、計数選別型抜き取り検査の方法を適用した S-SDV の方法を提案している。この方法では、治験実施計画で規定された被験者の来院（以下、Visit）ごとに収集したデータをサンプリングして SDV を行い、誤りを訂正したデータ数に基づき、各 Visit のデータに対する SDV の実施の有無を都度判断する。これにより、入力を誤る可能性が高い Visit 時のデータを重点的に SDV することが可能となった。しかし、Rutger

の方法では、被験者の Visit ごとに SDV の有無を都度判断する必要があることに加え、SDV を実施する Visit の判断が都度変更となるため、運用の複雑化が想定される。また、方法の設計背景やその動作特性は明示されておらず、標準的な方法として各製薬企業が導入するのは困難である。

大島 [3] は、データの重要度と訂正頻度を考慮したサンプリング方法を提案した。大島の手法では、過去の治験において訂正頻度が高かった CRF のページ（以下、Form）と、重要度が高い Form を SDV 対象とする。しかし、大島は領域が限られた 4 治験のみを分析対象としており、過去の治験におけるデータの訂正頻度の傾向が、S-SDV の適用対象とする治験においても同様な傾向であるのか、明確にしていない。そのため、過去の治験の結果を参照して SDV 対象とする Form を選定する方針の妥当性は不明確である。

2.2. 研究方法

本研究では、データの重要度と訂正頻度を考慮した S-SDV の方法を設計し、その際に考慮した事項を S-SDV の設計方法として手順化する。

まず、過去の治験の監査証跡データを用いたデータの訂正頻度の分析と、治験従事者を対象とした調査により、データの訂正頻度に影響を与える要因（以下、訂正要因）を検討する。つぎに、特定したデータの訂正要因をもとに、各治験のデータの訂正頻度を推測し、どのような治験に対して、積極的に S-SDV を適用すべきが明らかにする。そして、その結果をもとに、治験ごとに S-SDV の適用方針を検討し、それぞれ方法を具体化する。さらに、過去の治験の監査証跡データを用いて疑似的に本研究の提案手法を適用し、動作特性の検討および有用性の検証を行う。

さいごに、S-SDV の設計に際して検討した事項や、その検討背景を整理し、S-SDV の設計方法として提案する。

3. S-SDV の設計

3.1. 治験ごとのデータの訂正頻度の検討

データの訂正要因を特定し、治験ごとのデータの訂正頻度の傾向を検討するため、過去に実施された 18 治験の監査証跡データの分析結果から、18,037 件の訂正データを抽出した。なお、本研究では、2 回以上の入力処理が監査証跡データに記録されたデータのうち、SDV に代表される、製薬企業のモニタリング業務によって訂正されたと推測されるデータを訂正データと判断し、抽出対象としている。

治験、施設毎に、入力データの総数を分母、訂正件数を分子とし、データの訂正割合を算出した結果を表 1 に示す。

表 1 より、治験の領域ごとに訂正割合にばらつきがみられることがわかる。特に、領域ががんの治験では訂正割合が高くなっており、治験の実施段階（以下、Phase）が Phase

表 1. 治験・施設ごとの訂正割合

治験略号	領域	Phase	施設数	症例数	訂正割合	施設の訂正割合 max min
J3	がん	1	2	18	5.56	6.41 4.58
A1	がん	2	5	17	2.79	3.53 0.71
K1	がん	1	5	22	1.86	4.71 1.19
O7	がん	1	1	6	2.1	2.10
G5	疼痛(がん)	3	35	74	3.21	7.38 0.00
G6	疼痛(がん)	3	10	22	3.62	7.10 0.57
C1	高血圧	2	3	9	0.48	0.77 0.29
C6	高血圧	2	2	47	0.68	1.50 0.16
C4	糖尿病	2	70	892	2.02	5.13 0.29
C9	糖尿病	3	25	50	1.88	11.11 0.58
L3	糖尿病	2	25	453	0.7	1.59 0.30
L4	糖尿病	2	6	118	0.61	1.02 0.00
L5	糖尿病	2	1	24	0.02	0.02
M1	アルツハイマー	1	1	110	0.05	0.77 0.00
D9	インフルエンザ	3	10	20	0.24	0.05
C7	薬物相互作用	1	1	28	0.09	0.09
C8	薬物相互作用	1	1	37	0	0.00
C10	薬物相互作用	1	1	93	0.03	0.03

2, Phase 3 の治験も、訂正割合が比較的高くなっていった。一方、健常人を対象とした、領域ががん以外の Phase 1 の治験では、いずれの治験でも訂正割合が 0.1%以下となっており、非常に低い水準であることがわかった。

さらに、施設ごとの訂正割合の最大値と最小値には、ばらつきがみられ、治験の実施施設ごとに、訂正割合の水準が異なることが明らかになった。

3.2. 訂正要因と S-SDV の適用方針の検討

3.1 節で明らかにした治験の領域、施設ごとの訂正割合の傾向が、再現性のある傾向であるかを確認するため、特性要因図を作成して、データの訂正要因を複数検討した。

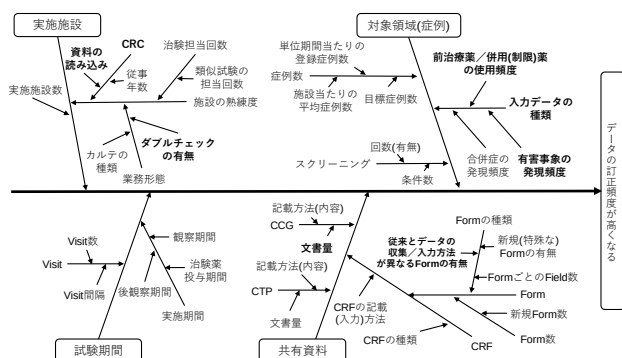


図 1. データの訂正要因

また、発生頻度の高い訂正要因を明らかにするため、医療機関で治験業務全般をサポートする治験コーディネーター (CRC) 16 名に対して、訂正要因の影響度に関する調査を実施した。図 1 で検討したデータの訂正要因のうち、影響度が高いと推測される訂正要因を計 16 個、評価項目として整理し、各訂正要因に対して 3 段階で影響度を評価した。その結果、「実施医療機関および CRC の熟練度」、「入力データの複雑性」、「CRF の構成の複雑性」に関する要因が、訂正頻度に大きく影響していることが明らかになった。

上記の訂正要因は、治験の領域、Phase ごとに特徴づけられる要素であり、例えば、領域ががん以外の Phase 1 の治験は健常成人を対象とするため、製薬企業は特定の施設に治験実施を依頼することが多い。そのため、「実施医療機関および CRC の熟練度」は高い場合が多く、また、健常成人を対象とするため「入力データの複雑性」は低い。したがって、訂正割合が低くなったと考えられる。

以上より、以下を S-SDV の適用方針として、適用方法の検討を行うこととした。

- データの訂正割合が比較的低い、領域ががん以外の Phase 1 の治験では、積極的に SDV 対象データを絞り込む。
- 領域ががんもしくは Phase 2, Phase 3 の治験には、治験の途中経過からデータの訂正割合を推定したうえで、SDV 対象データの絞り込み具合を施設ごとに判断する。

3.3. S-SDV の適用方法の検討

検査は、誤りの発生頻度、影響の大きさ、検知の容易さを考慮して設計を行うのが一般的である [4]。この中で、検知の容易さについては、SDV で誤りは容易に検出可能であるため、本研究では考慮しない。

誤りの発生頻度、すなわちデータの訂正割合は、過去の治験の情報から、あるいは、対象の治験の一部の SDV の結果から、推測する方法があり得る。

影響の大きさ、すなわち治験の結果解釈に与えるデータの重要度については、大島 [3] が CRF の Form 単位で 3 段階評価する方法を提案しており、本研究でもそれを用いる。治験結果の解釈に重大な影響を及ぼす評点 3 の Form のデータに対しては全数 SDV を行い、影響がわずかな評点 1 の Form のデータに対しては SDV の頻度を極力減らす。影響が中程度である評点 2 の Form のデータについては、データの訂正割合に基づいたサンプリング方法を検討する。

以上の点と、3.1 節で検討した治験ごとの S-SDV の適用方針を考慮し、具体的に S-SDV を設計した。

- 領域ががん以外の Phase 1 の治験は、原則、治験開始時から評点 3 の Form のみを SDV 対象とする。
- 領域ががん、もしくは Phase 2, Phase 3 の治験は、治験開始から一定期間は評点 2 および 3 の Form のデータを SDV 対象とし、特定のタイミングで評点 2 の Form のデータの訂正割合を、施設ごとに評価する。そのうえで、評点 2 の Form のデータに対する SDV の有無を施設ごとに決定する。

ここで、II 案については、訂正割合の「評価の基準」、および、「評価のタイミング」を検討する必要がある。

まず、「評価の基準」については、治験途中での評価となるため、評価後の訂正割合の変動を考慮する必要がある。そこで、訂正割合のロジット変換値に正規分布を仮定し、「訂正割合の信頼上限値」を指標とすることにした。区間推定を用いることにより、データの入力件数に応じた評価を行うことができ、治験初期でデータの入力件数が少なく、訂正割合が安定していない施設に対して厳しい評価を行うことが可能となる。なお、訂正割合の許容限界値は、製薬企業が治験の目的に合った信頼性、すなわち担保したいデータの品質に応じて決定すべきであり、本研究では、評点 2 の Form のデータが治験結果の解釈に重大な影響を及ぼさないことを考慮し、3%に設定することとした。

つぎに、「評価のタイミング」は、運用の難易度を考慮して、月毎に、全施設同一のタイミングで評価を実施することとした。また、評価の信頼性を担保するためには、評価時以降に訂正割合が上昇するリスクが低い時点で評価を実施する必要がある。そこで、過去に実施された 3 治験分 (表 1, A1, C4, O7) の監査証跡データを用いて、特定施設における月毎の訂正割合の信頼上限値の推移を調べた。

その結果、実施期間が 1 年を超える治験では、開始から 3 か月程度はデータの入力件数が少ない場合が多く、訂正

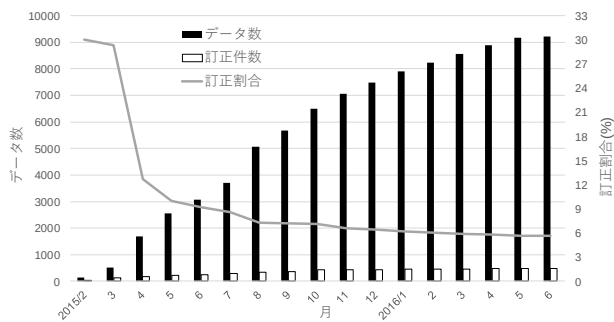


図 2. 訂正割合の月毎の推移(治験 C4 / 特定施設)

割合のばらつきが大きい傾向がみられた。そこで、訂正割合が大きく上昇するリスクが低くなる時点として、「初回データ入力日から4か月経過」した施設を評価対象とすることとした。そして、一定以上の施設が評価対象になると考えられる、「治験開始から6ヶ月経過時点、以後4ヶ月おき」に、全施設同時に評価を行うこととした。

3.4. 事例への適用と有用性の検証

3.3 節で決定した方法の動作特性と有用性を確認するため、過去に実施された4治験(表1, A1, C4, C9, C10)の監査証跡データを用いて、疑似的に方法を適用した。なお、前節で検討した方針に基づき、治験 A1, C4, C9 に対してはII案を、治験 C10 に対してはI案を適用することとした。

まず、製薬企業の治験担当者と Form のデータの重要度を評価した。治験 C4 では、評点1, 2, 3 の Form がそれぞれ、1, 12, 9 個、治験 C10 では、1, 12, 10 個となった。ここで、I案を適用した治験 C10 については、評点3 の Form のデータのみを SDV 対象とした。

つぎに、II案を適用した3治験については、訂正割合の評価時点を治験開始から6ヶ月経過時点と、そこから4ヶ月おきに設定した。そして、施設ごとに各評価時点での評点2 の Form のデータの訂正割合を算出し、その両側95%信頼上限値が3%未満の施設は、評点3 のデータのみを SDV の対象とすること(以下、合格)とした。治験 C4 に対する、手法の適用結果の概要を以下に示す。

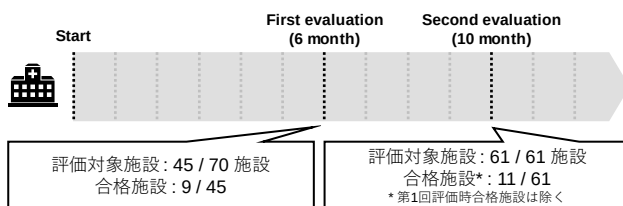


図 3. 適用結果の概要(治験 C4)

また、手法を評価するために、SDV 対象とならなかったデータの割合である「工数削減率」、SDV されず誤りが訂正されなかった評点2 のデータの割合である「未訂正率」を算出した。さらに、II案を適用した治験については、合格施設のうち、実際の治験終了時でのデータの訂正割合が許容限界値以上であった施設の割合である「誤判定率」も算出した。手法を適用した4治験の結果を、表2に示す。

表2より、「未訂正率」や「誤判定率」といった手法の信頼性を評価する指標に関しては、いずれの治験においても1%以下ときわめて低い値となっており、データの品質を十分に担保できているといえる。また、効率性を評価する「工

表 2. 有用性の評価

指標	算出式	C10	C4	C9	A1
工数削減率	非SDV対象データ数	59.47%	12.48%	19.34%	17.23%
	総データ数				
未訂正率	重要度2の未訂正データ数	0.02%	0.15%	0.07%	0.12%
	重要度2のデータ数				
誤判定率	合格のうち訂正割合が許容値以上の施設数	-	0.00%	0.00%	0.00%
	合格施設数				

数削減率」に関しては、全データを SDV の対象とする場合と比較して、I案では約60%程度、II案では12%から20%程度の削減率となっている。工数削減率は、データの重要度の評価や、設定する訂正割合の許容限界値に依存するため、一概に値の高低を判断するのは困難である。しかし、手法の信頼性を評価する、誤判定率と未訂正率の値を十分に保ったうえでの工数削減であることを考慮すると、十分に評価できる値と考えられる。

また、II案を適用した3治験において、訂正割合の許容限界値を変動させた際の信頼性、および効率性の推移を検討し、手法の動作特性を評価することとした。そこで、許容限界値を3%から5%まで、0.5%ずつ変動させ、その際の工数削減率と未訂正率をそれぞれ算出した。

表 3. 手法の動作特性の評価

治験	指標	訂正割合の許容限界値				
		3.0%	3.5%	4.0%	4.5%	5.0%
C4	工数削減率(%)	12.48	16.09	17.55	20.11	21.86
	未訂正率(%)	0.15	0.19	0.23	0.28	0.32
C9	工数削減率(%)	19.34	21.19	25.73	26.49	27.62
	未訂正率(%)	0.07	0.22	0.35	0.36	0.38
A1	工数削減率(%)	17.23	17.62	17.62	17.62	17.62
	未訂正率(%)	0.12	0.13	0.13	0.13	0.13

実施施設数が少ない治験 A1 では、許容限界値の変動が各指標に与える影響は小さいが、実施施設数が比較的多い治験 C4 と治験 C9 では、許容限界値が5%の場合、工数削減率は、3%の値と比較して、10%程度上昇している。なお、未訂正率はいずれの試験においても1%以下となっており、手法の信頼性は十分に担保できているといえる。

さらに、2.1 節で挙げた、Rutger [2]の手法と、大島 [3]の手法を比較対象とし、効率性(SDV 対象とするデータの削減度)、信頼性(誤ったデータを見逃がす頻度)、汎用性(適用可能治験の種類)、運用難易度(運用の煩雑性)の4つの観点から、提案手法の有用性を、◎、○、△、×の4段階で定性的に評価した。その結果を表4に示す。

表 4. 手法の有用性の評価

評価の観点	本研究		大島の手法	Rutgerの手法
	To Phase1	To Oncology, Phase2,3		
効率性	◎	△	○	○
信頼性	○	◎	△	○
汎用性	◎		×	◎
運用難易度	○		○	×

本研究の提案手法を評価すると、領域ががん以外の Phase 1 の治験に対しては、治験開始時から SDV 対象とするデータを重要データに絞り込むため、大幅に SDV の工数を削減可能である。しかし、あくまで過去の治験の結果に基づいた方針であるため、実施治験のデータの訂正傾向を逐次的に反映して SDV を実施するのは困難である。そのため、完全に信頼性を担保することは難しいといえる。

一方、領域ががん、もしくは Phase 2, Phase 3 の治験に対しては、重要度が中程度以上のデータを、一定期間 SDV する必要がある。そのため、SDV 対象とするデータの削減率は、他の手法と比較すると限定的である。しかし、治験実施施設のデータの訂正傾向に基づいて、SDV 対象とするデータを絞り込むため、信頼性は非常に高いといえる。

また、本研究の提案手法は、製薬企業 X 社において実際に稼働しており、汎用性、運用難易度の観点からも、高い評価を得られている。ただし、監査証跡データを逐次分析する必要があり、施設ごとに SDV 対象とするデータが異なる運用となるため、運用難易度はやや高いといえる。

上記のように、各手法の有用性を 4 つの観点から考察した結果、提案手法は、特に信頼性と汎用性の観点で、他手法よりも有用性が高くなっていることがわかった。以上より、本研究の提案方法は、データの品質を保ったうえで効率的に労力を削減できる、有用な方法であるといえる。

4. S-SDV の設計方法の提案

治験ごとに領域や、治験デザインは異なるため、3 章で提案した S-SDV を各製薬企業が直接的に導入することは困難である。また、本研究では、監査証跡データの分析によって明らかになった過去治験におけるデータの訂正傾向に基づき、データの重要度と訂正頻度を考慮した S-SDV を設計したが、データの重要度や訂正頻度の評価の粒度や、その判定基準の検討方法は多岐にわたると考えられる。そのため、S-SDV を設計する際には、本研究における検討事項や、検討の背景を踏まえて考案する必要がある。

そこで、上記の内容を、S-SDV の設計方法として手順化し、表 5 に整理した。また、本研究での検討結果とは異なる方針を採用するにあたっての提言も整理した。これにより、各製薬企業が、治験データの信頼性を担保する S-SDV を、効率的に設計することが可能になると考えられる。

5. 考察

本研究では、従来検討されていたデータの重要度のみでなく、データの信頼性保証において重要な要素である、データの訂正頻度にも着目した S-SDV の設計方法を提案した。単に S-SDV の手法やその適用結果を明示するだけでなく、設計の根幹となっている品質管理の考え方で併せて整理することで、S-SDV の設計や改善の効率化、および S-SDV の導入推進に寄与することが可能になった。

本研究の提案手法では、監査証跡データから、施設ごとにデータの訂正割合を算出し、SDV 対象とするデータの絞り込み具合を決定する。そのため、施設評価を実施した後に、何らかの要因によってデータの誤りが集中的に発生した場合等、未訂正率が高まってしまうリスクを否定できない。本研究で事例とした、製薬企業 X 社では、手法を適用し、各施設の優劣性を評価する際、監査証跡データから算出した訂正割合のみでなく、データの訂正に要した時間等、Key Risk Indicator と呼ばれる治験のリスクを計る指標の値や、現場モニターの意見を参考に、各施設への対応を逐次検討している。その点において、監査証跡データから得られた結果のみから判断するのではなく、どのような前提においてデータの訂正割合を算出しているかを理解し、S-SDV を運用することが重要といえる。

また、監査証跡データは、監査目的、すなわち治験が規制や計画書に従って実施され、その結果が正確に記録されたことを確認するために用いられる。そのため、監査目的以外で、このデータを用いた分析はされていなかった。しかし、本研究では監査証跡データから CRF 入力データの訂正頻度を把握し、S-SDV を実施する方法を提案している。その点において、本研究は監査証跡データの分析の有用性を明らかにしたといえる。

6. 結論と今後の課題

本研究では、監査証跡データを分析し、データの重要度と訂正頻度に基づいた、新たな S-SDV を提案した。

今後は、データの訂正内容を分析することで、治験の実施中に逐次プロセスを改善する方策の検討が求められる。

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”，部会資料
- [2] Rutger M. van den Bor,(2016) : “Efficient Source Data Verification Using Statistical Acceptance Sampling: A Simulation Study”，Therapeutic Innovation& Regulatory Science, Vol.50, No.1, pp.82-90
- [3] 大島一真ら(2016) : “監査証跡データに基づくサンプリング SDV に関する研究”，日本品質管理学会第 46 回年次大会研究発表会，pp.119-122
- [4] 日本品質管理学会(2009) : 「新版 品質保証ガイドブック」，一般社団法人日本品質管理学会，pp.504-507

表 5. S-SDV の設計方法(一部)

検討事項【大項目】	検討事項【小項目】	検討結果	検討背景	設計に向けた提言
データの品質管理方針の決定	SDV対象データ選定時の観点の検討	データの「重要度」と「訂正頻度」	・製造業の信頼性検査における考慮事項を参考 ▶重要度・不良率・検出難易度 ・SDVでは誤りの検出難易度がどのデータも同様と仮定	・重要度のみを考慮する方針も考えられるが、検査設計では製品の不良率(データの訂正頻度)を考慮するのが一般的 ・データの訂正頻度を把握するために、監査証跡データが有用 ・RBMの中で検討されているKPIをもとに、SDVの度合を決定する方針も考えられる ・ただし、KRIと訂正頻度の相関は要検討
	評価の粒度(単位)の決定	Formごと	・データのForm/Fieldごとに重要度が異なる ・Formの中のデータ項目(Field)ごとに重要度が異なる場合もあり ・しかしField単位だと評価プロセスが煩雑化	・Fieldごとに重要度がばらつくFormは、CRFの設計段階でFormを分割することも可能 ・ただし、CRFの品質低下リスクは考慮すべき ▶観察/検査項目単位での分割でなくとも、誤入力を誘発してしまう可能性有 ・Fieldごとに評価を行うことで、より精緻な評価が可能
	重要度の評価方法の検討	3段階評価	・実際の運用を考えると、「訂正頻度によらずSDV対象とする重要データ」「訂正頻度に応じてSDVの是非を検討するデータ」「訂正頻度によらずSDV対象としないう非重要データ」の3段階での評価が妥当	・5段階評価等詳細な基準を作成し、重要度に応じて許容する訂正頻度を変動させる方法も有効(重要度5: 必ずSDV対象、重要度4: 訂正割合1%以下の場合SDV非対称、重要度3: 訂正割合3%以下の場合SDV非対称、重要度2: 訂正割合1%以下の場合SDV非対称、重要度1: 必ずSDV非対象、等)
	評価の基準の決定	評価基準を策定	・治験担当者に重要度評価を依頼(試験目的に応じて治験ごとに評価) ・複数治験の評価結果/評価の根拠を蓄積	・評価結果を蓄積/共有していくことで、精緻な評価基準の策定が可能 ・重要度は複数部門の治験担当者で議論したうえで評価すべき ・本研究で検討した評価基準も参考になる ▶重要度3: 有害事象、被験者背景、被験者番号、等 ▶重要度2: 併用薬、併用制薬原、合併症、等 ▶重要度1: 臨床検査、薬力学測定用採血、等(外部データ)
訂正頻度の評価方法の検討	評価の粒度(単位)の決定	治験分類ごとに訂正頻度の評価方法を策定	・治験のPhaseと領域によってデータの訂正頻度に傾向有 ▶データの訂正頻度に影響を与える要因を検討し、傾向の再現性を確認	・過去治験の監査証跡データを分析し、治験ごとのデータの訂正割合の傾向を捉えたうえで分類を検討すべき ・データの訂正要因を観点とし、各治験のデータの訂正頻度を推定することも可能
...	...	...	...	...