# *RBQM in CDM -*
## *Getting ready for targeted querying?*

**November 2021**

**Peter Stokman,**
**Bayer**

**CDM Leadership Network**

# RBQM in CDM - getting ready for targeted querying?

**Content**

1. The quality of the Data Collection process *per se*, and the Value of the back-end Queries
2. What Regulators have been encouraging us to do, how Transcelerate helped, & what we (Sponsor companies, Service Providers) have been implementing
3. The new Data Collection Process and the potential to significantly improve CDM's efficiency

Encore: very quick overview of what we have been developing at Bayer

# RBQM in CDM - getting ready for targeted querying?

1. The Quality of the Current Data Collection Process *per se*

and

the Value of the back-end Queries

# RBQM in CDM - getting ready for targeted querying?

## CDM Network: An Inquiry into Queries and Quality

Journal of the Society for Clinical Data Management

Stokman P, et al. Risk-based Quality Management in CDM. *Journal of the Society for Clinical Data Management*. 2020; 1(1): 1, pp.1–8.
DOI: https://doi.org/10.47912/jscdm.20

ORIGINAL RESEARCH

### Risk-based Quality Management in CDM

An inquiry into the value of generalized query-based data cleaning

Peter Stokman*, Lisa Ensign†, Dirk Langeneckhardt‡, Monika Mörsch§, Kristin Nuyens‖, Diego Herrera¶, Gernot Hochgräber*, Valérie Cassan**, Philip Beineke†, Richard Kwock†, Annemarie Voortman‖, Sven Vogelgesang‡, Sami Boussetta†† and Bradley Bitzer*

**INTRODUCTION:** The essence of Risk-based Quality Management is to focus on what matters most. Sponsor companies are annually spending many billions of dollars on automated and manual data cleaning, utilizing internal and site resources to create the illusion of an error-free data base. However, over the last 15 years, a robust data acquisition process was developed, and much of the current query-based data cleaning approach is a self-imposed remnant of the error-prone processes of the past.
**OBJECTIVE:** Motivated by previously reported insights regarding the limited value of generalized source data verification (SDV), seven pharmaceutical companies and a clinical trials solution vendor teamed up to answer the question "How efficient is the query-based data cleaning process?"

Galápagos

Pierre Fabre

BAYER

Almirall

SANOFI PASTEUR

Boehringer Ingelheim

MERCK

medidata

Stokman et al., Risk-based Quality Management in CDM. Journal of the Society for Clinical Data Management. 2020; 1(1): 1, pp. 1–8.

# RBQM in CDM - getting ready for targeted querying?

## RBQM in Clinical Development: Monitoring



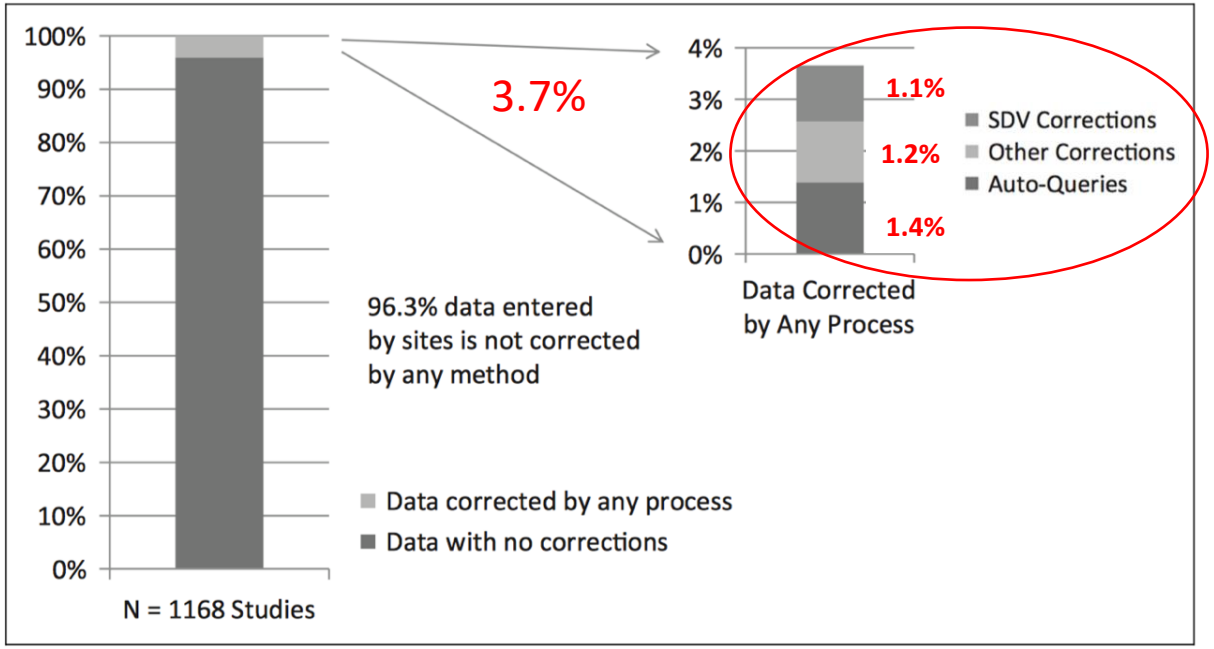



**Figure 3.** Percentage of electronic case report form corrections attributed to source data verification (SDV) versus other data correction methods.

Evaluating Source Data Verification as a Quality Control Measure in Clinical Trials Nicole Sheetz, Therapeutic Innovation & Regulatory Science, 2014, Vol. 48(6) 671-680

# RBQM in CDM - getting ready for targeted querying?

## RBQM in Clinical Development: Monitoring

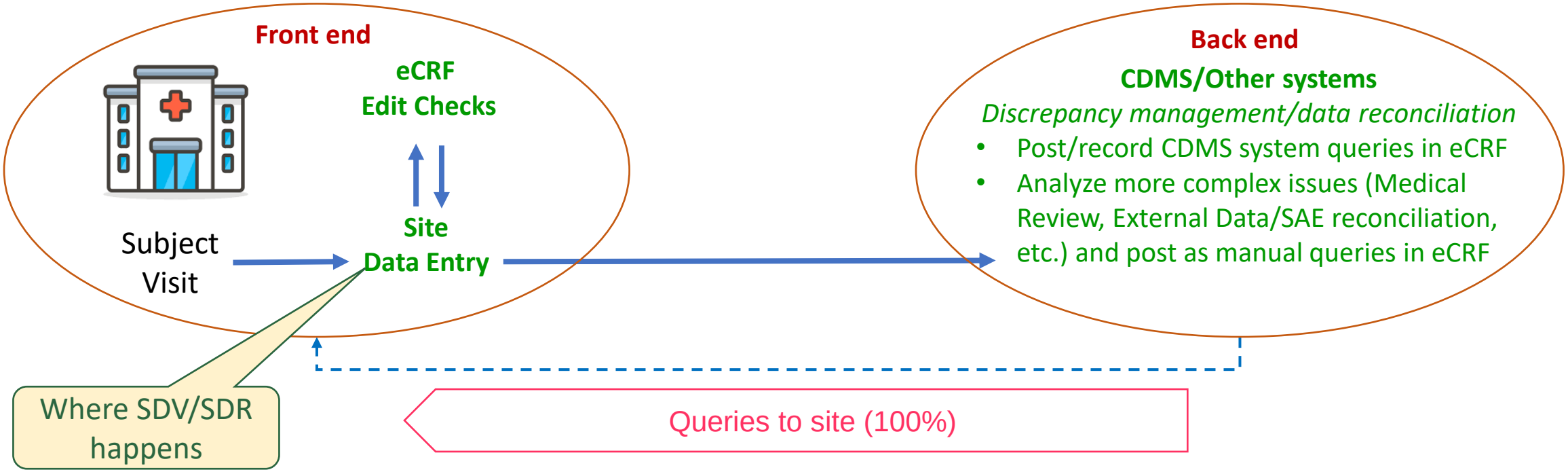
**The importance of Sheetz et al., 2014**

- Generalized SDV results in just 1.1% of data being updated
- This finding lead Sponsors and CROs to adopting targeted SDV approaches, guided by
  - Site performance indicators (including ‘Changes due to SDV’)
  - Data Criticality
- Targeted SDV still corrects transcription errors where needed
- Risk-based approach to monitoring - shifting from SDV to more strategic approaches like
  - Source Data Review (SDR)
  - Central Statistical Monitoring

# RBQM in CDM - getting ready for targeted querying?

## The Clinical Data Management Process, late 2010s

Very High Level



# RBQM in CDM - getting ready for targeted querying?

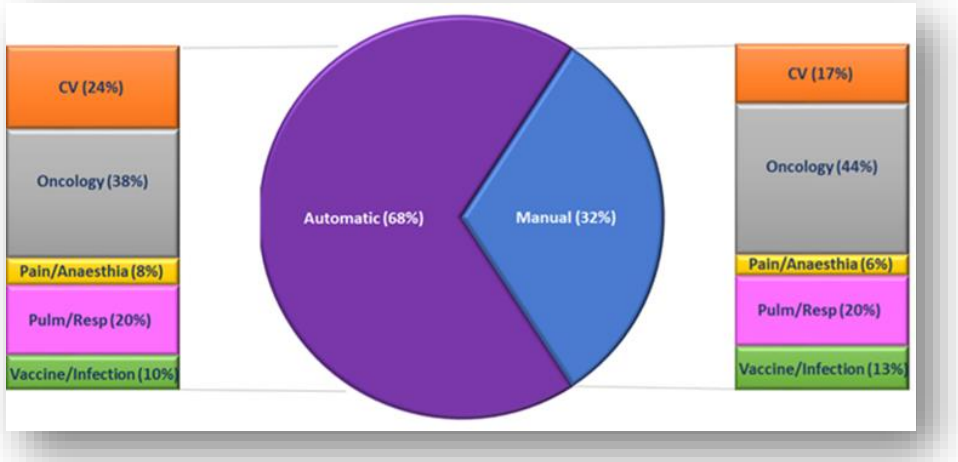
## Why is it relevant to look at the value of the back-end query process?

- Query-based data cleaning consumes resources from sponsor and clinical site.
  - Back-end queries: specified, programmed and tested
  - Manual back-end queries: articulated by the Clinical Data Manager, Study Medical Expert, Drug Safety Specialist.
  - Automatic & Manual queries: reviewed and responded to by site personnel.
- The cost of queries: $28 to $225 per query
- No guidance stipulating the details of the query management process ⇒ largely self-imposed
- *100%* query resolution is at odds with a risk-based (i.e., targeted) approach to quality management

⇒ An inquiry into the efficacy of this process is justified

# RBQM in CDM - getting ready for targeted querying?

## Data Points & Queries




**Methods**

- Twenty Phase III studies, 5 TAs,
- completion dates between 2014-2019, randomly selected

**Result**

- 49,259,945 data points
- 1,939,606 queries were generated
- 1,327,526 (68%) automatically triggered
- 612,080 (32%) manually created

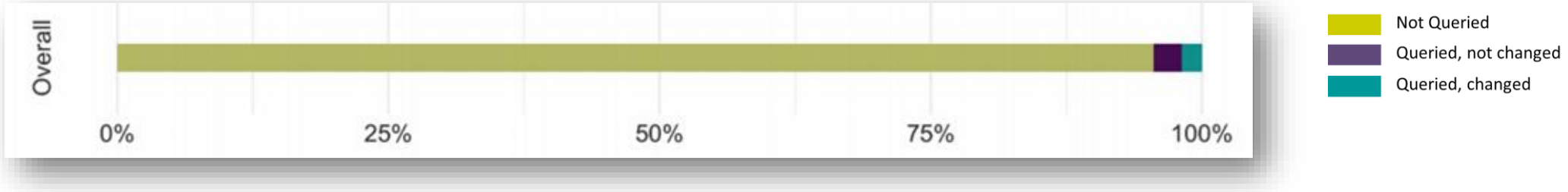
- Query rate = 3.9% (number of queries/total data points)

- On average 96,980 queries per Phase 3 study (@ $28-$225/query => $3M to $22M/study)

# RBQM in CDM - getting ready for targeted querying?

How many data are actually changed? Or: what is the added value of all of this?

- 3.9% of all the data collected is ever queried
- On average, these queries lead to a data change in 42%
- In 49,259,945 datapoints, Automatic + Manual queries ⇒ 818,921 changes = 1.7%
  - Automatic queries: 421,339 changes = 0.9% of datapoints
  - Manual queries: 397,582 changes = 0.8% of datapoints
- > 98.3% of data collected in a study is never changed as a result of the query process.



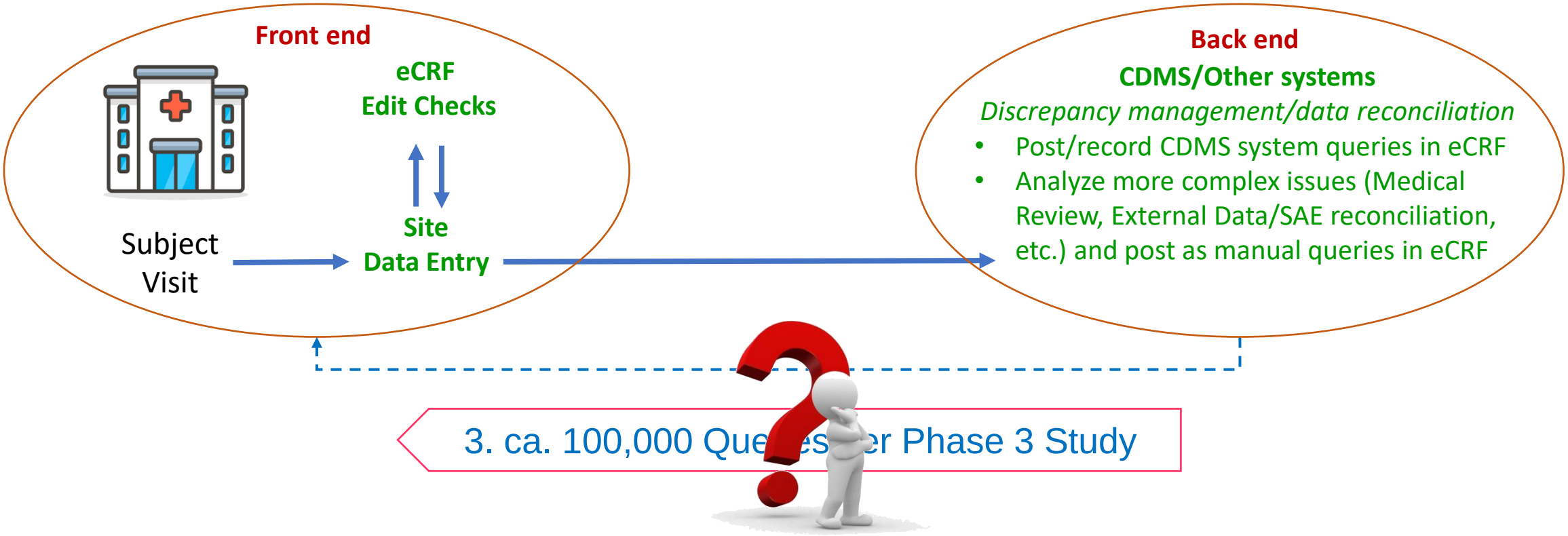

- About 2-3 % of the data changes are non-informative data changes
  e.g., change of format “Unknown Strength” corrected to blank, blank corrected to “Unknown”, or “UN:UN” corrected to “Unknown”

# RBQM in CDM - getting ready for targeted querying?

## Conclusions & Recommendations

1. This part of the process leads to a data accuracy of > 98%

2. This part of the process questions about 4% of the data & corrects about 50% of that
⇒ Changes <1.7%



# RBQM in CDM - getting ready for targeted querying?

2. What Regulators have been encouraging us to do,
how Transcelerate helped,
what we (Sponsor companies, Service Providers) have been implementing

# RBQM in CDM - getting ready for targeted querying?

## Guidance to encourage better results by using smarter approaches: focus on Monitoring

- ICH Q9 Quality Risk Management (January 2006): '… provides principles (…) for QRM that can be applied.'
- MHLW: Basic Principles of Risk-based Monitoring, Evaluation and Licensing Division, Pharmaceutical and Food Safety Bureau (July 2013)
- FDA Guidance:  Oversight of Clinical Investigations — A Risk-Based Approach to Monitoring (Aug 2013)
- Transcelerate: Position Paper: Risk-Based Monitoring Methodology (May 2013 + Updates)
- EMA Reflection Paper on risk-based quality management in clinical trials (November 2013)
- ICH E6 (R2) Addendum (June 2017)
  - ICH E6 (R3): 'ICH E6 (R2) included a focus on a proportionate, risk-based approach to the design and conduct of clinical trials. E6(R3) will be designed to further advance this concept and to encourage relevant parties to utilize this approach.'
- ICH E8 (R1) Designing Quality in Clinical Trials (May 2019)

  3.3.4: Reviewing Critical to Quality Factors: Build on accumulated experience and knowledge with periodic review of critical to quality factors to determine whether adjustments to risk control mechanisms are needed, since new or unanticipated issues may arise once the study has begun.

# RBQM in CDM - getting ready for targeted querying?

Transcelerate: RBQM, Indicators & Thresholds. For RBM; generalized to clinical development/CDM

RBQM: ensuring subject safety and data quality by:

1. building QbD into trials
2. early and ongoing risk assessment,
3. a focus on Critical Processes and Critical Data,
4. use of Risk Indicators and Thresholds*, and
5. adjustment of activities based on the issues and risks identified throughout the study.

*Risk Indicators and Thresholds

1. Risk Indicators should be to be assigned with Thresholds => trigger an action
2. Risk Indicators, Thresholds and actions: documented in Integrated Quality Risk Management Plan
3. developed and finalized in a timely manner (i.e. during study planning, => including study-specific risks),
4. cross-functional collaboration.
5. collect feedback on Risk Indictor performance

# RBQM in CDM - getting ready for targeted querying?

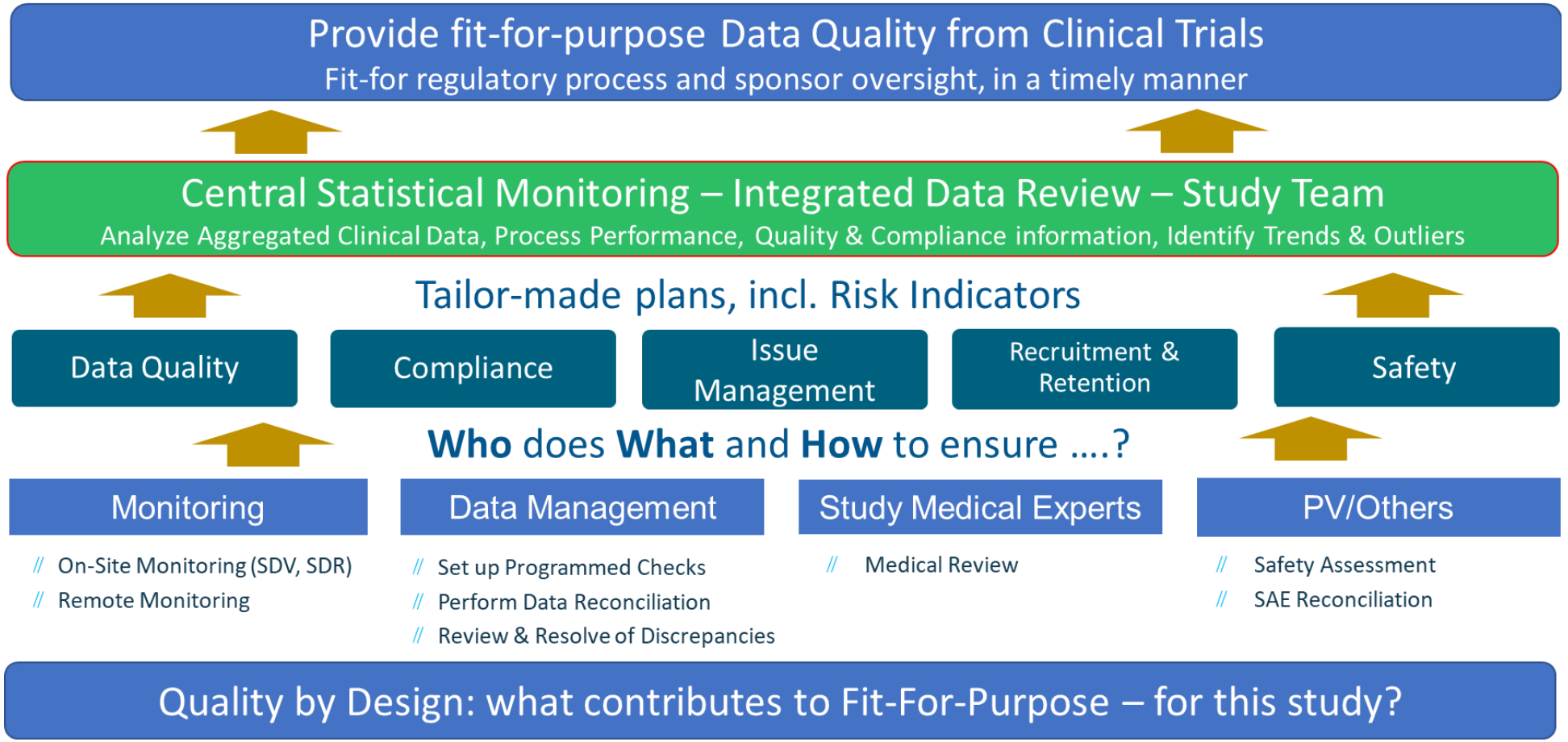
## Transcelerate Risk Indicator Library

- Blinded survey on Risk Indicators and associated benefits
- => Risk Indicator Library: 140 Risk Indicators
- RBM-focused - but cross-functional

| Risk Indicator Categories & Sub-Categories | | | | | | | |
|---|---|---|---|---|---|---|---|
| Monitoring | | | | CDM/Central Statistical Monitoring | | | |
| Investigational Product | Essential Documents | Staffing, Facilities and Supplies | CRA/On-site Workload | Data Quality | Issue Management | Safety | Subject Recruitment and Discontinuation |
| CRA compliance | Responsibility for Data Quality | PI Oversight | CRA Compliance | Adverse Events | CRA compliance | Adverse Events | Enrollment |
| Dispensing compliance | Site Compliance | Site Compliance | CRF Review | CRF Completion | Discrepancy (Query) Management | Lab Data | Screen Failures |
| Drug Accountability | Study compliance | Site/Staff Turnover | | Data Trends | Protocol Compliance | Outcome or Endpoint Issues | Subject Discontinuation |
| Patient compliance | | | | Diagnostic Testing | Site Compliance | Serious Adverse | |
| Shipment | | | | Discrepancy (Query) Management | Subject Discontinuation | Study Drug Discontinuation | |
| Site compliance | | | | Patient Compliance | | | |
| Storage | | | | PI Oversight | | | |
| | | | | Protocol Compliance | | | |

# RBQM in CDM - getting ready for targeted querying?

## RBQM is a cross-functional endeavor



# RBQM in CDM - getting ready for targeted querying?

Examples of Indicators. Analyze @ Study, Country, Site level; potential QTLs in Red

Compliance

- **Compliance**
  Protocol (PDs/VFs, In/Ex)
- **Timeliness**
  Subject visit, Sampling, Data Entry, SAE reporting, Adjudication
- **Recruitment & Retention**
  Enrolment, Screen Failure rate/reason, Discontinuation Rates/Reasons, LFU

Safety

- **AE/SAEs**
  Incidence, Frequency, AEs of special interest, Discontinuations, Causality assessment

Data Quality & Issue Management

- **Transcription Errors**
  Changes due to SDV
- **Turn Around Times**
  Data Entry, Query aging, Issue resolution
- **Queries**
  Counts (per site/CRF), Re-queries, Query Rate, Query Quality, Changes due to Queries
- **Missing Pages/Data**
  Identification
- **Numerical & Categorical Data**
  Variance, Outliers, Anomalies
- **Advanced Analytical approaches**
  Duplicate randomization, AE underreporting
  Unsupervised approaches, Bots, ML/AI

# RBQM in CDM - getting ready for targeted querying?

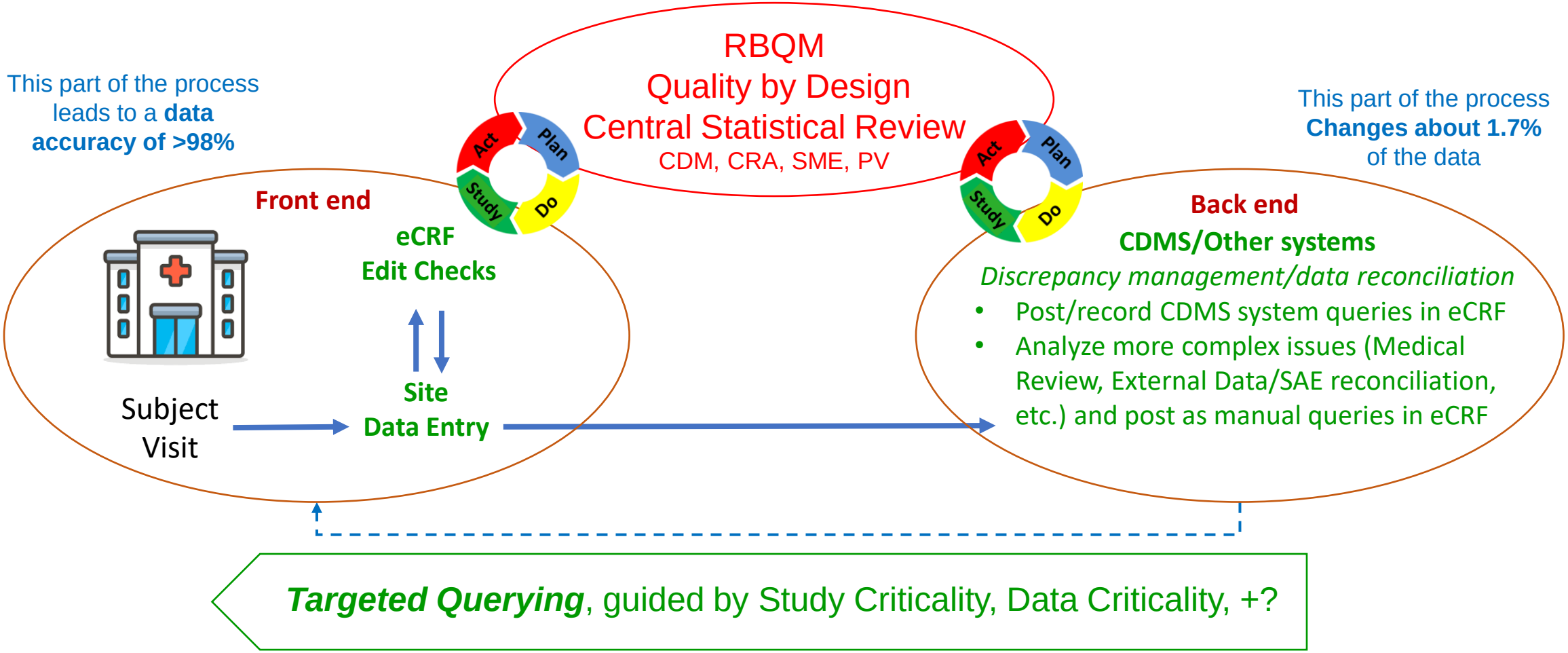
3. The new situation in CDM

and

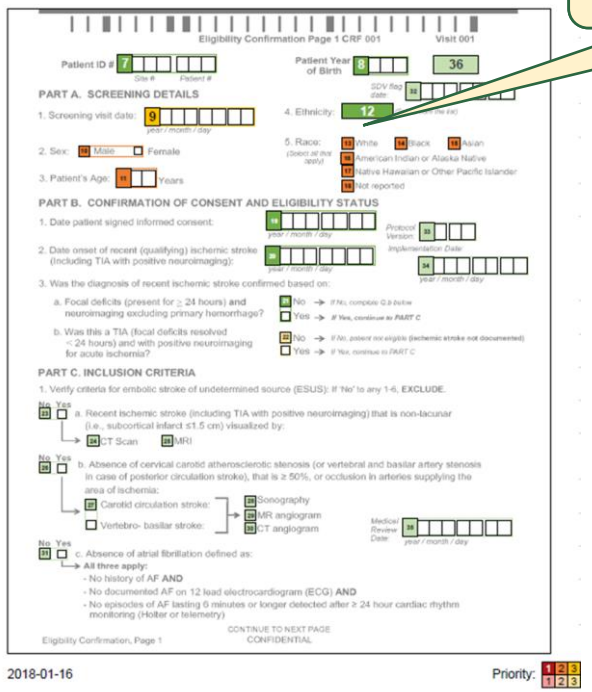
the potential to significantly improve our efficiency

# RBQM in CDM - getting ready for targeted querying?

## The New Clinical Data Management Process, early 2020s



# RBQM in CDM - getting ready for targeted querying?


Eligibility Confirmation Page 1 CRF 001 Visit 001

Patient ID # 7 Patient Year of Birth 8 36

PART A. SCREENING DETAILS

1. Screening visit date: 9 year / month / day

4. Ethnicity: 12

5. Race: 13 White 14 Black 15 Asian 16 American Indian or Alaska Native 17 Native Hawaiian or Other Pacific Islander 18 Not reported

2. Sex: 19 Male Female

3. Patient's Age: 11 Years

PART B. CONFIRMATION OF CONSENT AND ELIGIBILITY STATUS

1. Date patient signed informed consent: 16 year / month / day Protocol Version 33

2. Date onset of recent (qualifying) ischemic stroke (including TIA with positive neuroimaging): 26 year / month / day Implementation Date: 34 year / month / day

3. Was the diagnosis of recent ischemic stroke confirmed based on:

a. Focal deficits (present for ≥ 24 hours) and neuroimaging excluding primary hemorrhage? No → If No, complete Q.b below Yes → If Yes, continue to PART C

b. Was this a TIA (focal deficits resolved < 24 hours) and with positive neuroimaging for acute ischemia? No → If No, patient not eligible (ischemic stroke not documented) Yes → If Yes, continue to PART C

PART C. INCLUSION CRITERIA

1. Verify criteria for embolic stroke of undetermined source (ESUS): If 'No' to any 1-6, EXCLUDE.

No Yes 20 a. Recent ischemic stroke (including TIA with positive neuroimaging) that is non-lacunar (i.e., subcortical infarct ≤1.5 cm) visualized by: 24 CT Scan 25 MRI

No Yes 21 b. Absence of cervical carotid atherosclerotic stenosis (or vertebral and basilar artery stenosis in case of posterior circulation stroke), that is ≥ 50%, or occlusion in arteries supplying the area of ischemia: 27 Carotid circulation stroke: 28 Sonography 29 MR angiogram 30 CT angiogram Vertebro- basilar stroke: Medical Review Date 38 year / month / day

No Yes 31 c. Absence of atrial fibrillation defined as: All three apply: - No history of AF AND - No documented AF on 12 lead electrocardiogram (ECG) AND - No episodes of AF lasting 6 minutes or longer detected after ≥ 24 hour cardiac rhythm monitoring (Holter or telemetry)

Eligibility Confirmation, Page 1 CONTINUE TO NEXT PAGE CONFIDENTIAL

2018-01-16 Priority: 1 2 3

**Data Criticality**
- Determined per variable

**Study Criticality**

*Targeted Querying*

| | | Study Criticality | | | | | |
|---|---|---|---|---|---|---|---|
| | | Critical Studies | | | Non-Critical Studies | | |
| | | *Site Data Risk Level* | | | *Site Data Risk Level* | | |
| Data Criticality | Critical Data Points | High | Medium | Low | High | Medium | Low |
| | Data Point A | | | | | | |
| | Data Point B | | | | | | |
| | | | | | | | |
| | Non- Critical Data Points | High | Medium | Low | High | Medium | Low |
| | Data Point C | | | | | | |
| | Data Point D | | | | | | |
| | | | | | | | |

- Review & Query everything, Review & Query nothing, Review & Query a percentage?
- More dimensions: add site performance?

# RBQM in CDM - getting ready for targeted querying?

**Conclusions:**

1. Current data collection process is very robust, with an accuracy of over 98%
2. CDM's efforts to resolve all queries means a significant burden to the Sites - with a very small added value
3. Implementation of RBQM in Clinical Development led to the adoption of Quality by Design, and a suit of indicator dashboards and analytical tools
4. Enabling cross-functional, timely and prospective risk control & mitigation
5. New CDM context: the back-end query process may be ready to evolve into *targeted* querying – guided by data criticality
6. Potential for significant increase in CDM's efficiency – with a matching reduction of the burden on the sites

# RBQM in CDM - getting ready for targeted querying?

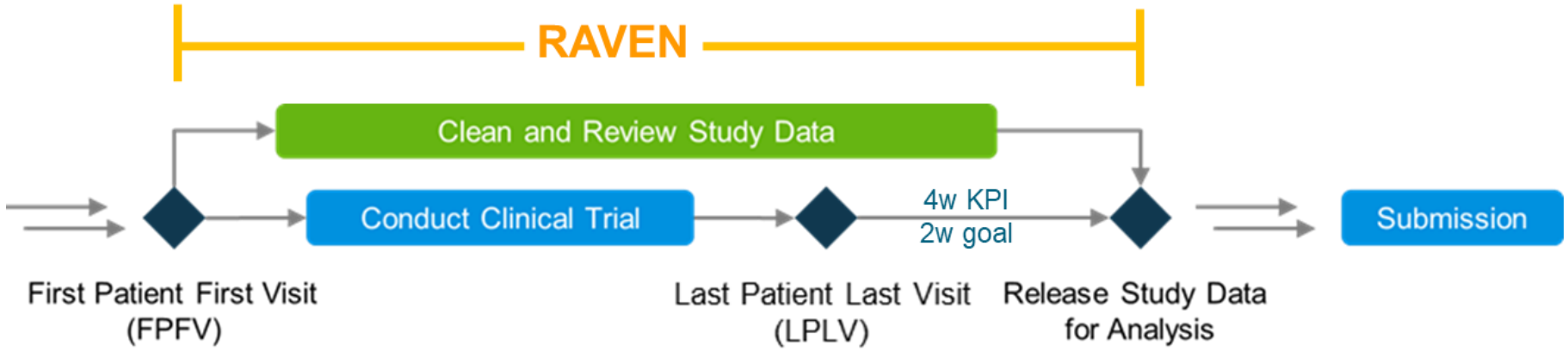
**RBQM in CDM: more than targeted querying - Status & Outlook**

1. 'RBQM in CDM': work in progress
   - Scope (eCRF, external data, eCOA data, ATR, …)
   - Minimal requirements (re Indicators, QTLs, …)
   - Relevant dimensions (Study criticality, Data criticality, Site Performance, …)
   - Process
   - Documentation
   - …
2. Working with the SCDM Innovations committee to develop guidance
3. Hope to also include suggestions from Regulators on RBQM in CDM

# RBQM at Bayer

**RAVEN in November 2021**

- RAVEN: a GCP-compliant Reporting And ReView ENvironment for *ongoing clinical trials*
- RAVEN concept is to “build dashboard once” and then use for all future studies
- RAVEN Scope: support review of the data flow:
  - Correctness, completeness, consistency, timeliness, process performance, progress tracking
  - Centralized Data Review: data quality, compliance, trends, issues, outliers, anomalies (IDRP)
  - Enables a risk-based approach as it visualizes where/when the risks occur => adapt, mitigate risks => Quality ↑, Efficiency ↑



# RBQM at Bayer

**RAVEN in November 2021**

- 85+ dashboards
  + 3 ‘Libraries’ (10 - 15 DBs each): Phase 1 Medical Review, DM Oversight, PharmacoVigilance;
- 800 Users
  - Clinical Data Managers (in-house & outsourced)
  - Central Statistical Monitoring
  - Phase 1 Medical Review Library for Study Medical Experts
  - Management: Study Health dashboards
  - PharmacoVigilance Library
  - Country & Study Lead Monitors, Study Managers, Clinical Research Associates

RAVEN Dashboards

A RAVEN ‘Library’