EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

HMA/EMA Joint Task Force on big data – Surveys results

Big Data Meeting
15 December 2017

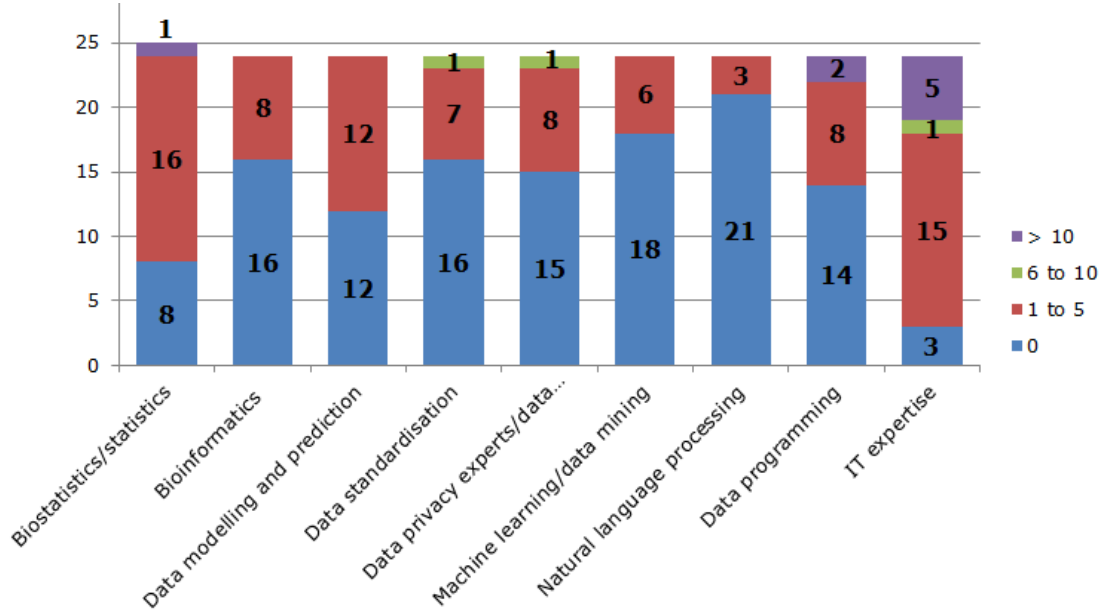
Presented by: Alison Cave, Aldana Rosso and Marjon Pasmooij

Background

- One of the key actions of the HMA/EMA Joint Task Force on Big Data, as set out in its mandate, is to carry out a **survey** to ascertain the current situation across the European regulatory network with regard to the available expertise and competences in the analysis and interpretation of Big Data.
- 2 surveys were launched one addressed to NCAs and one to pharma:
  - **NCAs which responded** (14th September to 9th October): **24** (Germany PEI, BVL and BfArM, Latvia, Denmark, Ireland, Malta, Slovak Republic, Finland, Norway, Czech Republic, Iceland, Hungary, Spain, Austria, Cyprus, Romania, Croatia, Netherlands, Slovenia, Belgium, Portugal, United Kingdom, Italy)
  - **Member States (EEA) which did not respond**: **9** (Bulgaria, Estonia, France, Greece, Liechtenstein, Lithuania, Luxembourg, Poland, Sweden)
  - **Pharma Industry** (20th September to 20th October): **37 responses**

NCA survey results

Q2 - What is your NCA's current in-house level of expertise relevant to big data analytics? Please indicate the approximate numbers of Full Time Equivalents (FTEs) with expertise in the following analytical areas



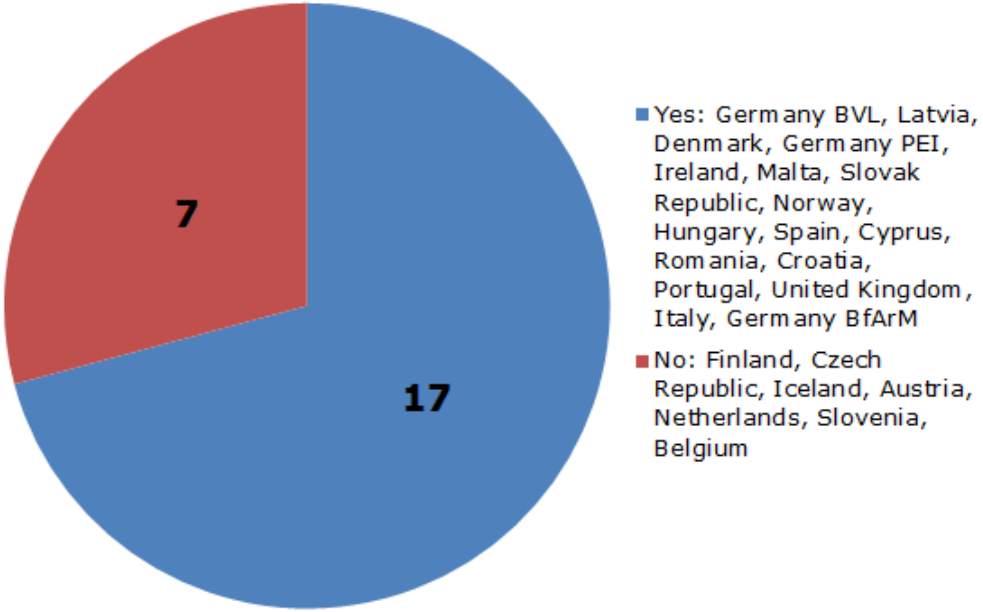
Main messages:

- **Majority of NCAs do not currently have FTEs** with specific expertise in big data analytics;
- Surprising results for (bio)statistics.

Notes:

- "IT expertise" possibly misunderstood as General IT support;
- Some of the expertise areas listed are not "in-house" but within data centres (e.g. SK, NL, UK), so are not core business for some NCAs.

Q3 - Does your NCA consider it necessary to increase the number of resources in house with expertise to analyse big data?



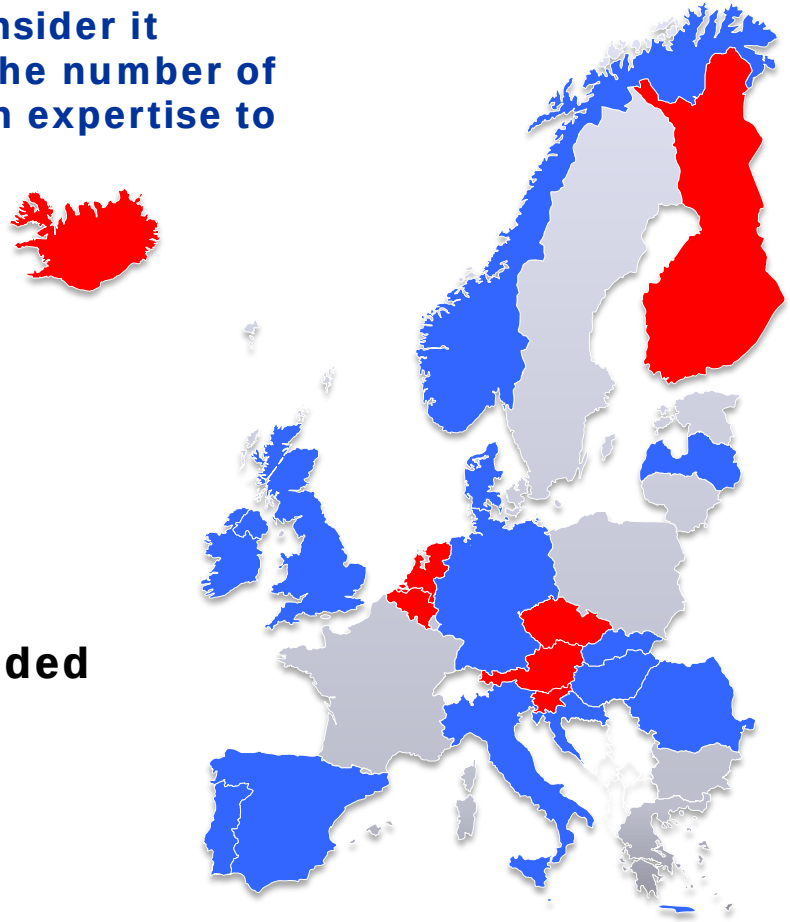
"Yes":

- **DE**: *terms big data is too broad, need for a more focussed discussion for medicines regulation;*
- **HR, UK**: *Use of big data can be extremely useful for various aspects of medicines regulation.*

"No":

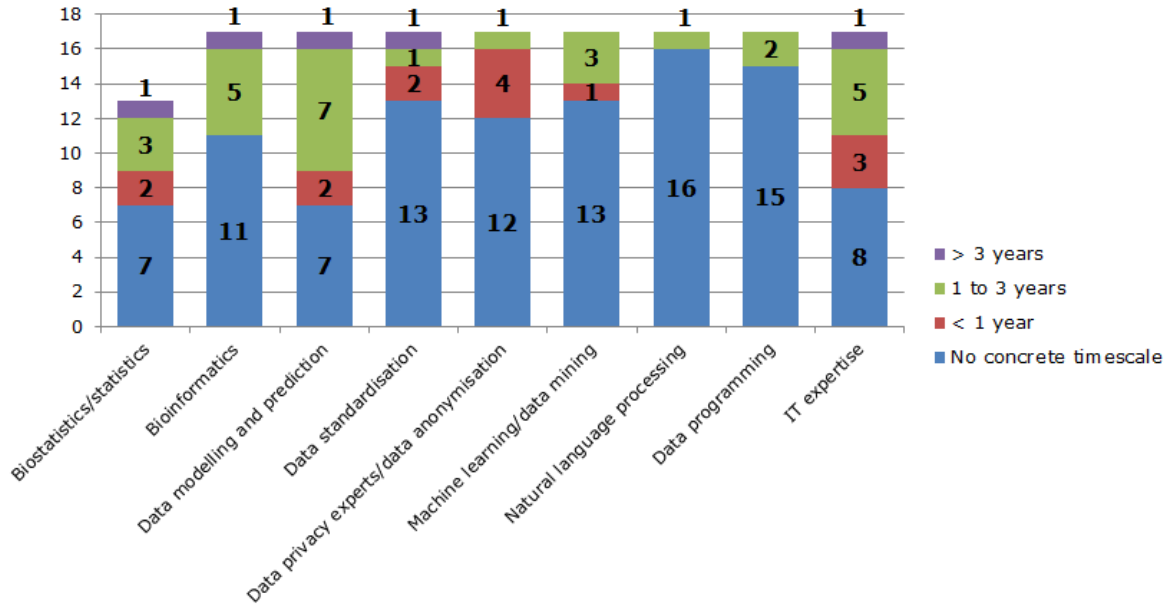
- **IS, SI**: *We currently do not handle big data, no sufficient data;*
- **NL**: *Expertise on big data but not a priority for FTE increase;*
- **BE**: *A "big data" expert team can be set up with the currently existing profiles in the agency.*

Q3 - Does your NCA consider it necessary to increase the number of resources in house with expertise to analyse big data?



Q3 - Does your NCA consider it necessary to increase the number of resources in house with expertise to analyse big data?

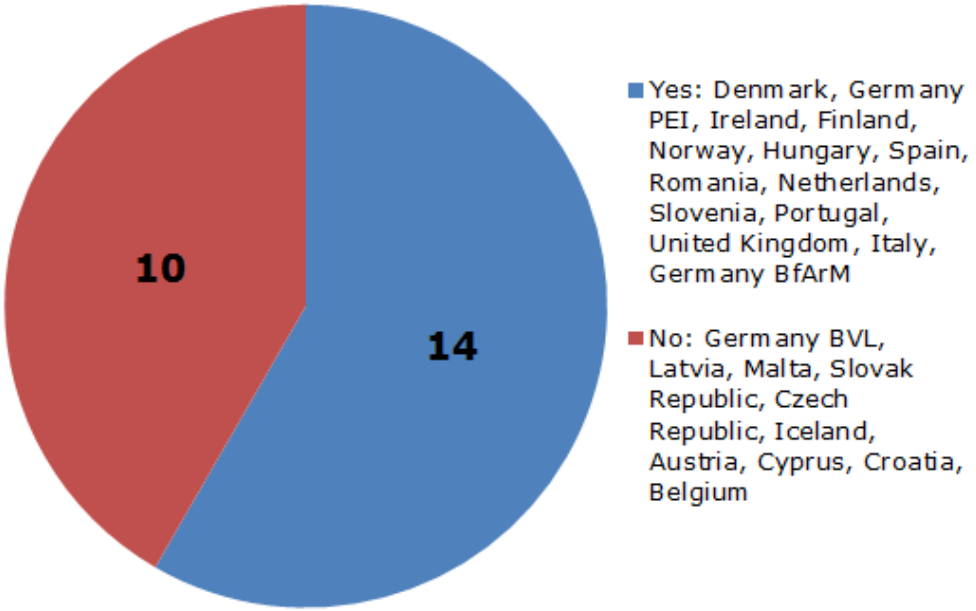
Timescales for the **16 NCAs** which responded "Yes":



Main messages:

- **2 / 3** of the NCA respondents consider an increase in FTEs with expertise in big data **necessary**;
- **But "No concrete timescale" for most** of the expertise areas apart from (bio)statistics, data modelling / prediction and IT expertise;
- The specified timescales are mainly **within 3 years.**

Q4 - Does your NCA have direct access to external big data sets to inform your decision making process?



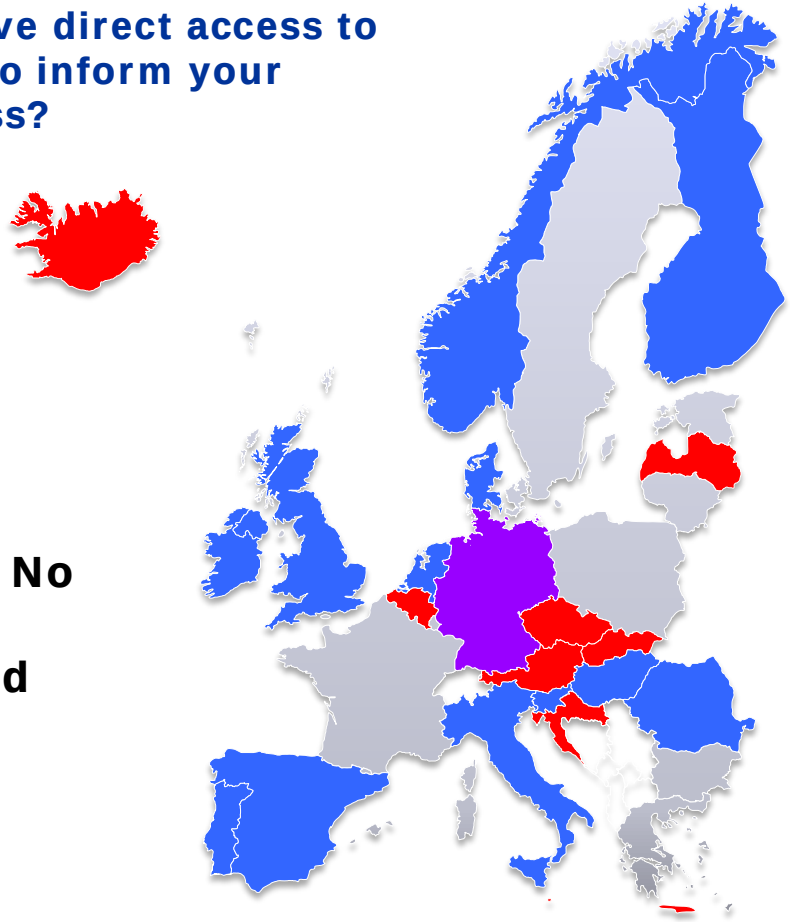
"<u>Yes</u>":

- **UK:** *Links with groups Genome England and biobanks, EHR frequently accessed. Social media data have been used in an ongoing IMI project but not yet in decision making.*

"<u>No</u>":

- **AT:** *no direct access to ext. datasets, no need for our decision making process. But we buy from ext. providers, e.g. consumption data of human antibiotics.*

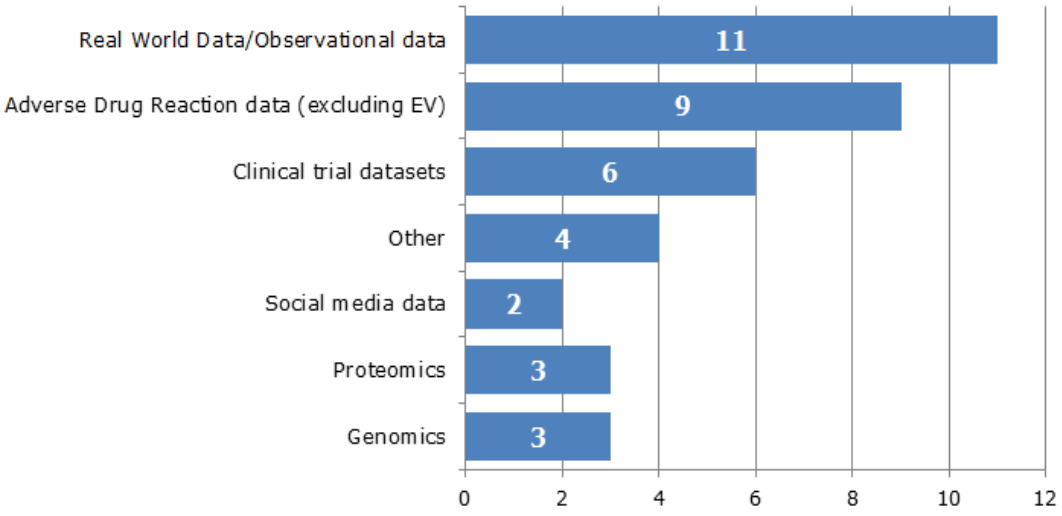
Q4 - Does your NCA have direct access to external big data sets to inform your decision making process?



- Yes
- No
- Both Yes and No
- Not responded

Q4 - Does your NCA have direct access to external big data sets to inform your decision making process?

Timescales for the **14 NCAs** which responded "Yes":



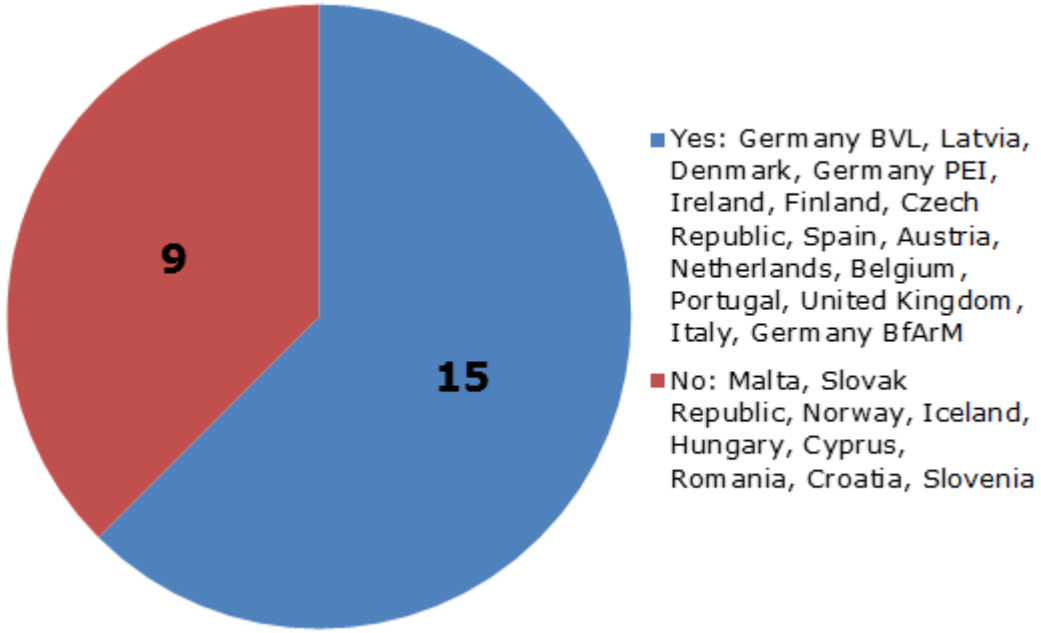
Main messages:

- **More than half** of the NCA respondents have **direct access** to external big data sets;
- **Mostly RWD / Observational data, ADR data and CT data**;

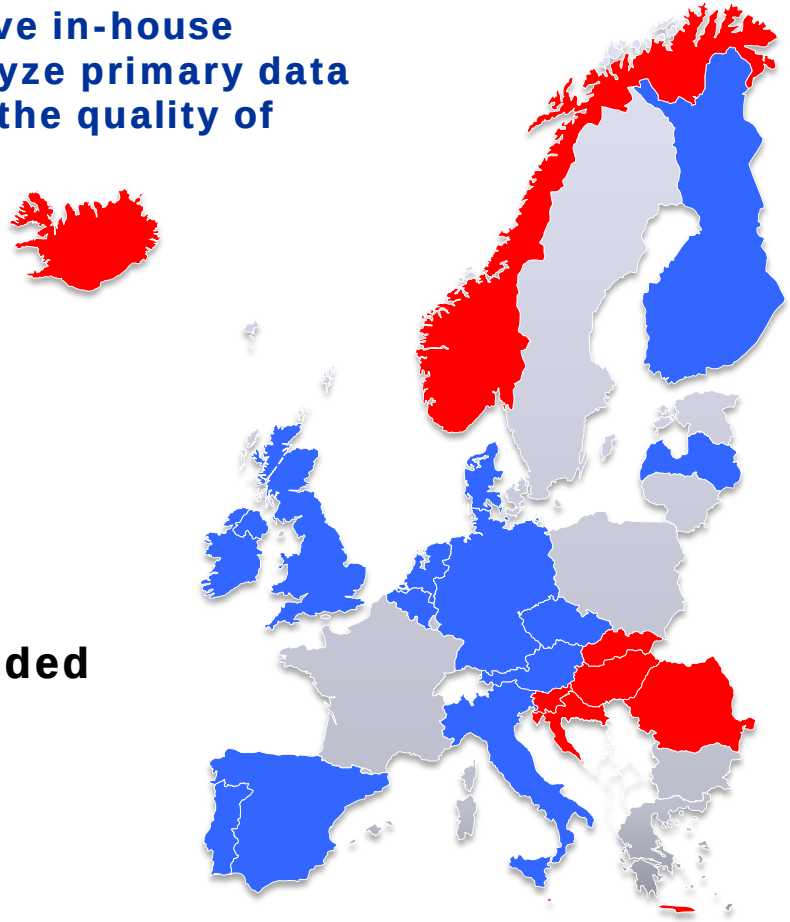
"Other":

- **HU:** National Health Insurance data base, ClinTrials.gov, EUDRA CT, EUDRA GMP, EnCEPP;
- **SI:** Pharmaceutical products consumption data;
- **PT:** Prescription and dispensing data.

Q5 - Does your NCA have in-house systems / tools to analyze primary data sets in order to assess the quality of submitted data sets?

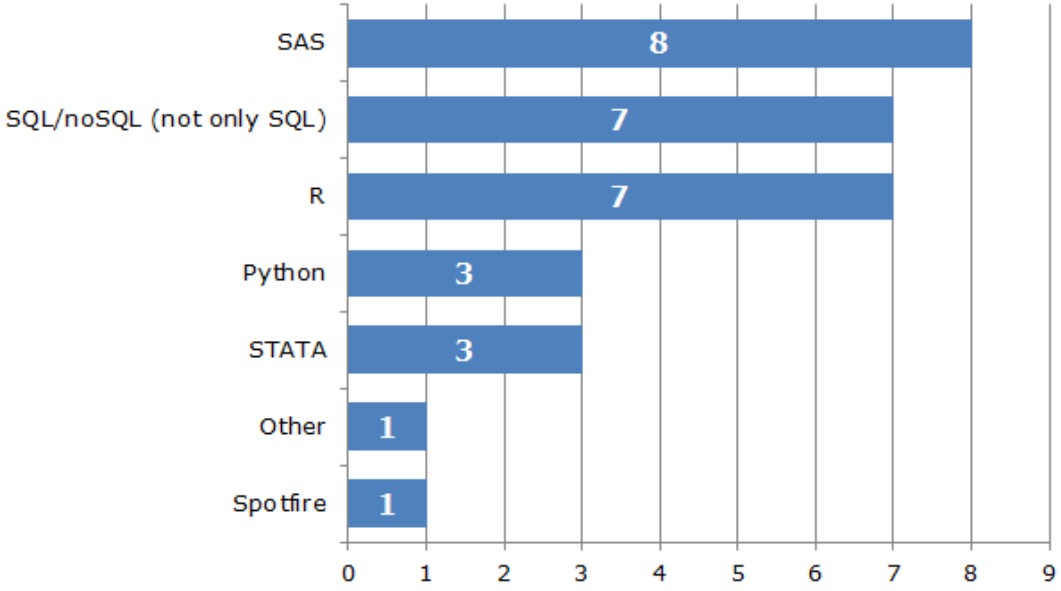


Q5 - Does your NCA have in-house systems / tools to analyze primary data sets in order to assess the quality of submitted data sets?



Q5 - Does your NCA have in-house systems / tools to analyze primary data sets in order to assess the quality of submitted data sets?

Timescales for the **15 NCAs** which responded "Yes":



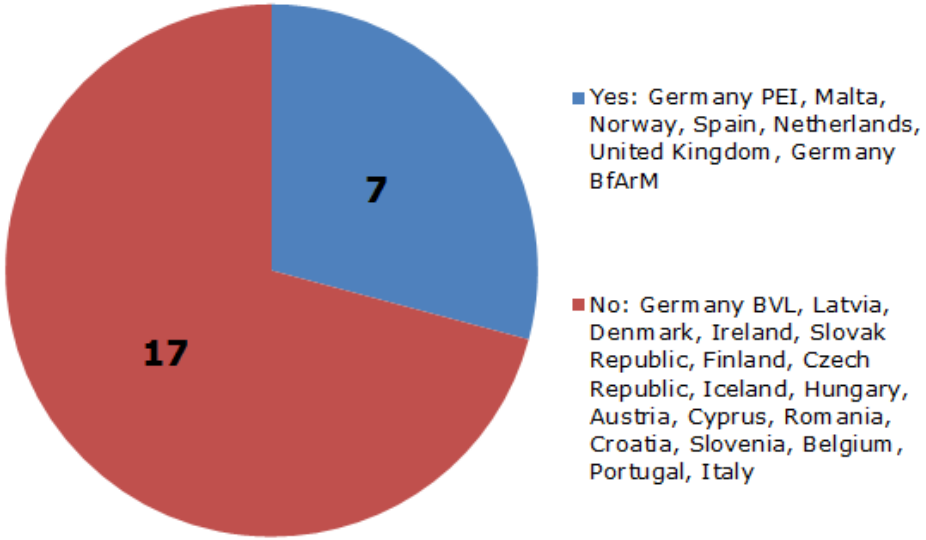
Main messages:

- **More than half** of the NCA respondents have in-house systems / tools to analyse primary data sets;
- **Mostly SAS, SQL, R**.

Notes:

- **"Other" - NL:** SPSS;
- **LV:** Oracle Discovery
- **AT:** *we use these tools for statistics, analysis and data validation, but don't consider this in the meaning "big data";*
- **UK:** *These are primarily used for the conduct of new research and not currently for assessing the quality of submitted data sets.*

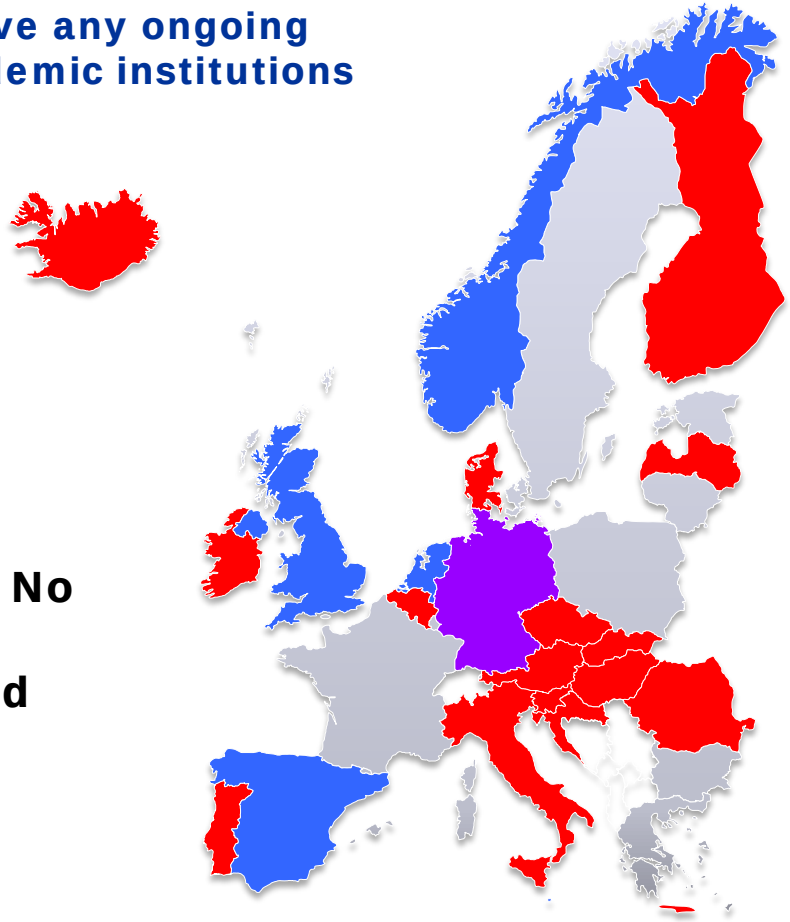
Q6 - Does your NCA have any ongoing collaboration with academic institutions within big data?



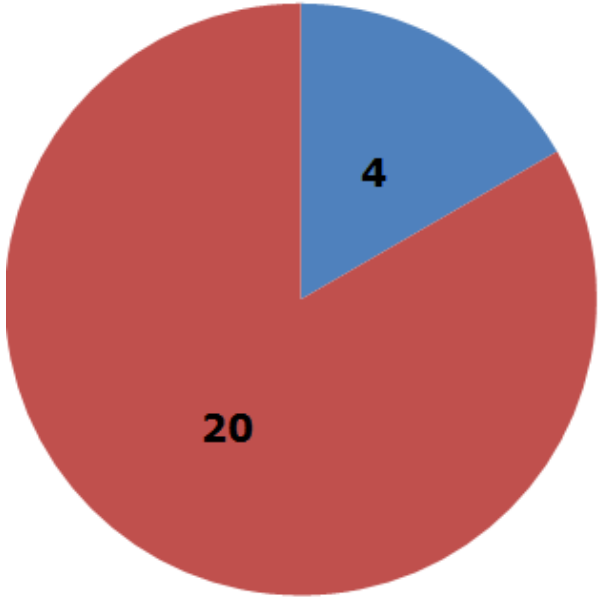
"<u>Yes</u>":

- **DE (PEI):** Other German agencies (DIMDI, RKI) hosting health insurance data; additionally scientific networks (IVDK, ESSCA, Nora: European registry) and various universities;
- **DE (BfArM):** Cooperation with research groups hosting or having access to big data, recent research projects involving health claims data with pmv forschungsgrupppe Cologne for drug safety questions;
- **MT:** University of Malta;
- **NO:** Cancer registry Norway;
- **ES:** University groups mainly for pharmacoepidemiology purposes and with computer engineering for natural language processing;
- **NL:** University of Rotterdam and Utrecht (esp for oncolytica);
- **UK:** Several UK universities/private companies/larger collaborations for analysis of EHR data, IMI WEBRADR (inc. universities/ MAHs / SMEs) for analysis of social media data, sit on a number of registry scientific advisory committees.

Q6 - Does your NCA have any ongoing collaboration with academic institutions within big data?



Q7 - Has your NCA outsourced any big data-related services?

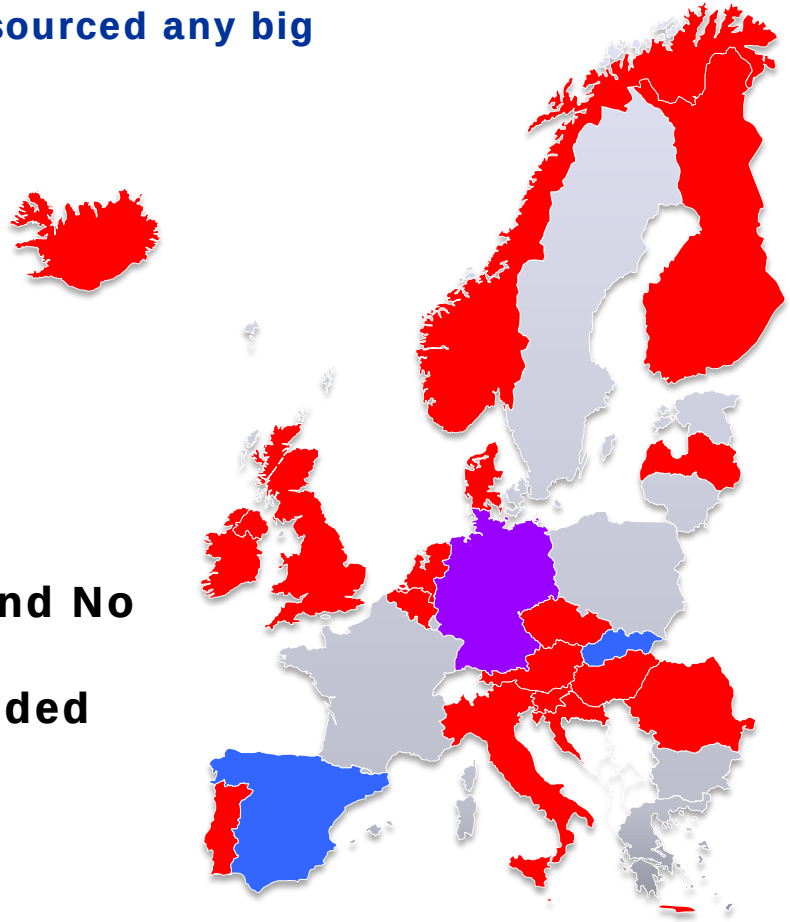


- Yes: Germany PEI, Slovak Republic, Spain, Germany BfArM
- No: Germany BVL, Latvia, Denmark, Ireland, Malta, Finland, Norway, Czech Republic, Iceland, Hungary, Austria, Cyprus, Romania, Croatia, Netherlands, Slovenia, Belgium, Portugal, United Kingdom, Italy

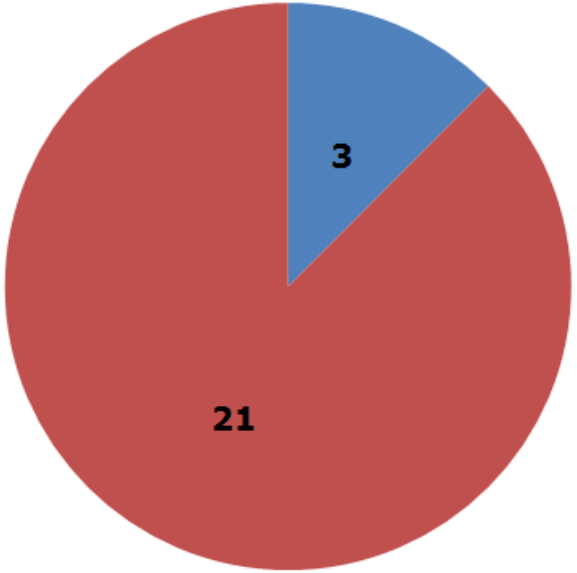
"Yes":

- **DE (PEI):** Commercial sequencing services;
- **DE (BfArM):** funding of research project in drug safety;
- **SK:** generation of drug consumption, data about re-export;
- **ES:** Computer engineering services for natural language processing and graph representation.

Q7 - Has your NCA outsourced any big data-related services?

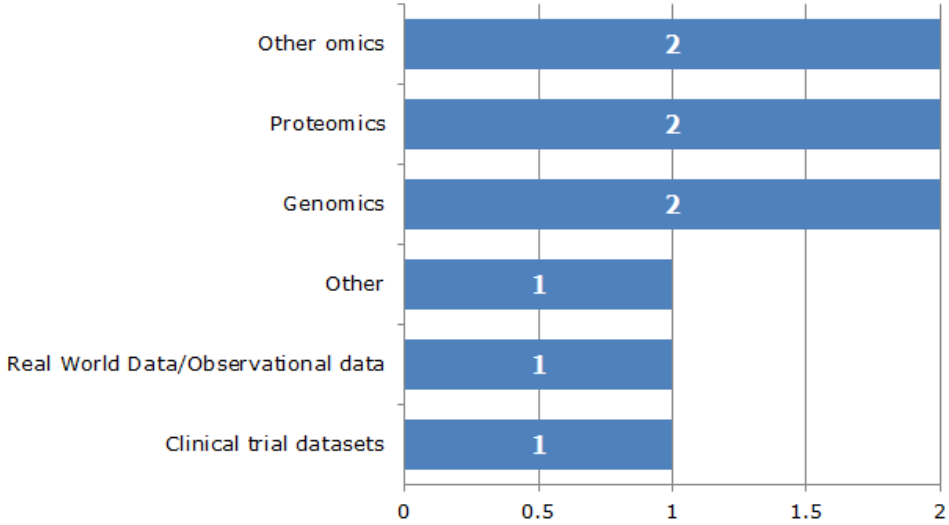


Q8 - Has your NCA received any requests from any sponsors asking for scientific advice at national level on the applicability or the analysis of big data sets?



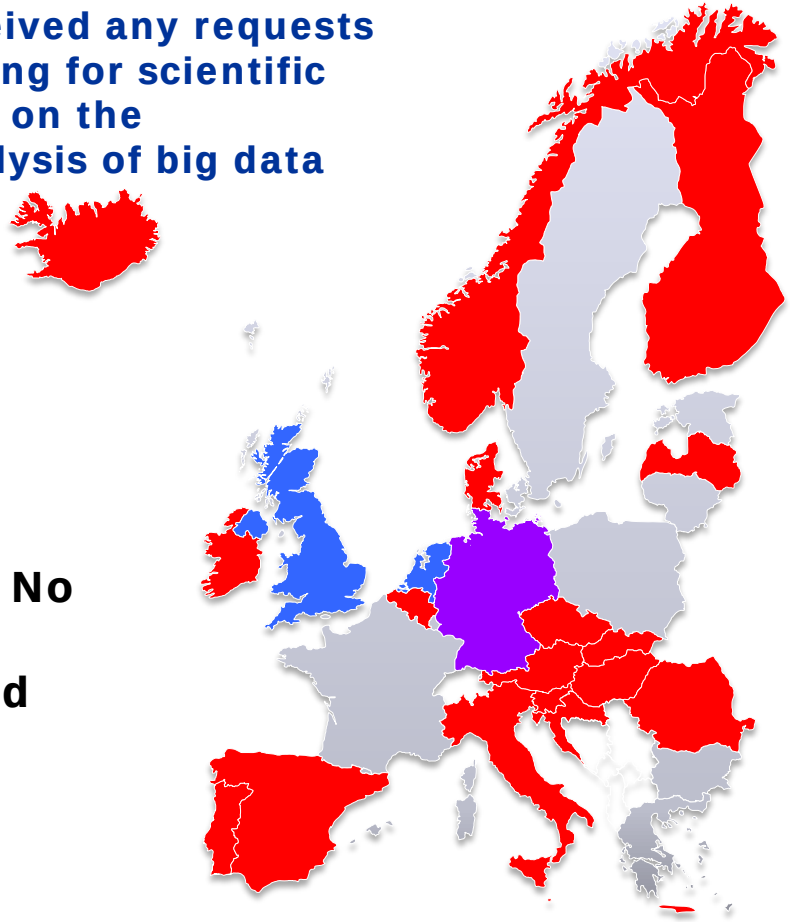
Yes: Germany PEI, Netherlands, United Kingdom

No: Germany BVL, Latvia, Denmark, Ireland, Malta, Slovak Republic, Finland, Norway, Czech Republic, Iceland, Hungary, Spain, Austria, Cyprus, Romania, Croatia, Slovenia, Belgium, Portugal, Italy, Germany BfArM

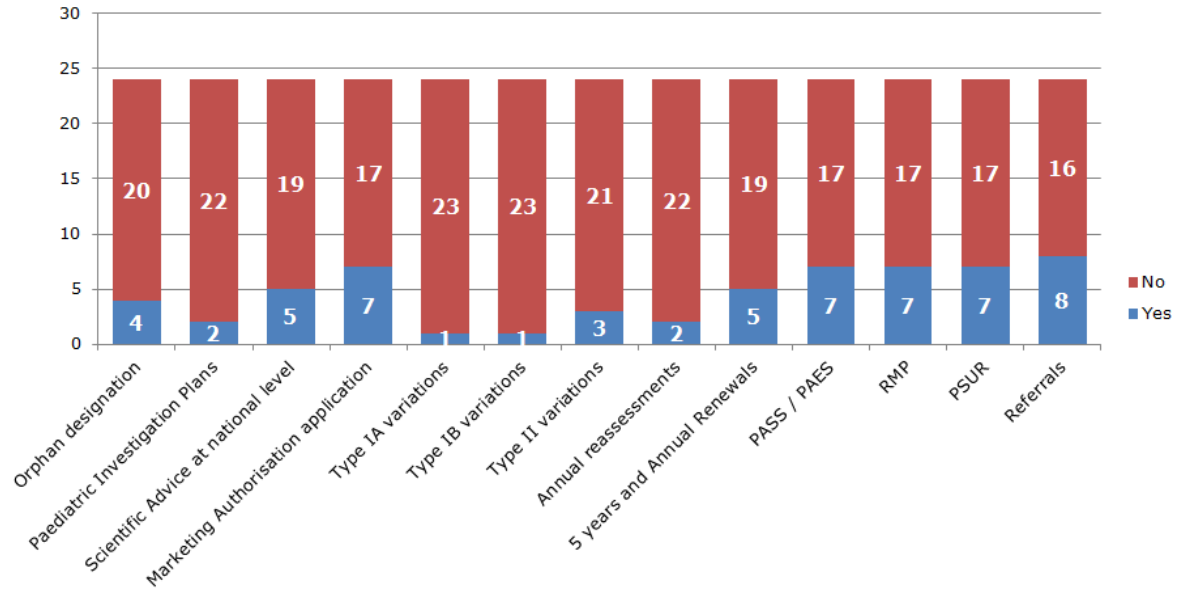


- **Germany PEI, Netherlands:** 1-5 requests received (e.g. from NL: question on data from a cohort - effect in a control group and with regards post approval studies);
- **United Kingdom:** 10-15 requests received.

Q8 - Has your NCA received any requests from any sponsors asking for scientific advice at national level on the applicability or the analysis of big data sets?



Q9 - Please indicate where big data has been used to derive evidence to support regulatory procedures, e.g. to understand disease progression, to better stratify patient populations, to identify an underlying mechanism of action for an ADR?



Big data mainly not used

Examples, challenges and comments provided:

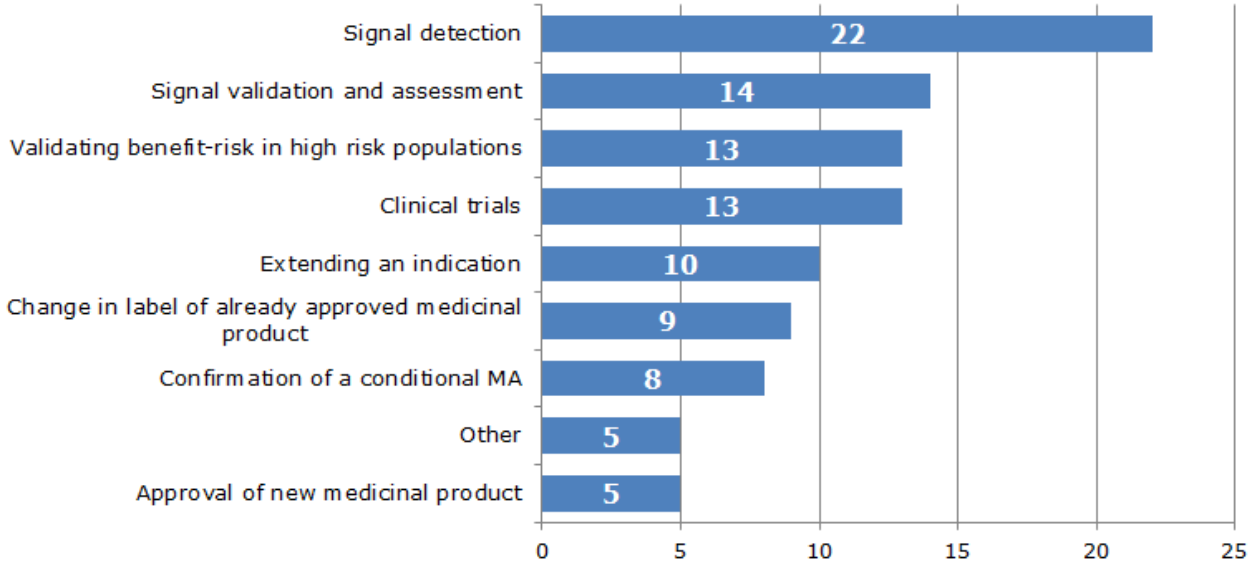
- **DE (BVL, PEI):** challenges for MAA: data management, data protection law in DE, timeliness of analysis;
- **DE BfArM:** Annual Reassessments for orphan drugs where registries collecting data are condition of MA. Data from health claims is often subject to safety studies in the setting of rare events diseases;
- **ES:** challenges for EHR or patient registries studies: feasibility of the study due to incompleteness of data or limitations of the data source;
- **RO, PT:** HTA assessment;
- **UK:** Big data featured in all procedures listed and also to support pharmacovigilance and larger national reviews;
- **IT:** These kind of analysis are not performed by AIFA but by the MAHs.

Q10 - Please provide specific examples, if any, where big data sources have been used to derive evidence to support label change decisions

- **6 NCAs** commented:

  - **DE (PEI):** *Big data sources used have so far not resulted in label changes but have been used to confirm safety of medicinal products;*
  - **DE (BfArM):** *COC Referral - decisions based on studies using health claims data; SmPC information on dosing for Cystadane based on post-marketing data from registry. Information on SCARs occurring with use of Epoetins from spontaneous reports from ADR databases*
  - **FI:** *Extrapolation of indications in biosimilars and drugs used in certain paediatric populations*;
  - **ES:** *Information of CV profile of aceclofenac*;
  - **UK:** *CPRD data to generate data in-house for the assessment of the safety of pertussis vaccine in pregnancy as part of a National vaccination campaign. This data was used to support removing the label warning against use in pregnancy*;
  - **IT:** *Art.31 Referral from 2013 requesting MAHs to review all available clinical/non clinical data including safety assessment of PV databases. A DUS was imposed to evaluate knowledge & behaviour of health professional of the adopted RMMs*.

Q11 - In which regulatory areas does your NCA see big data being first applied? Please select the relevant areas

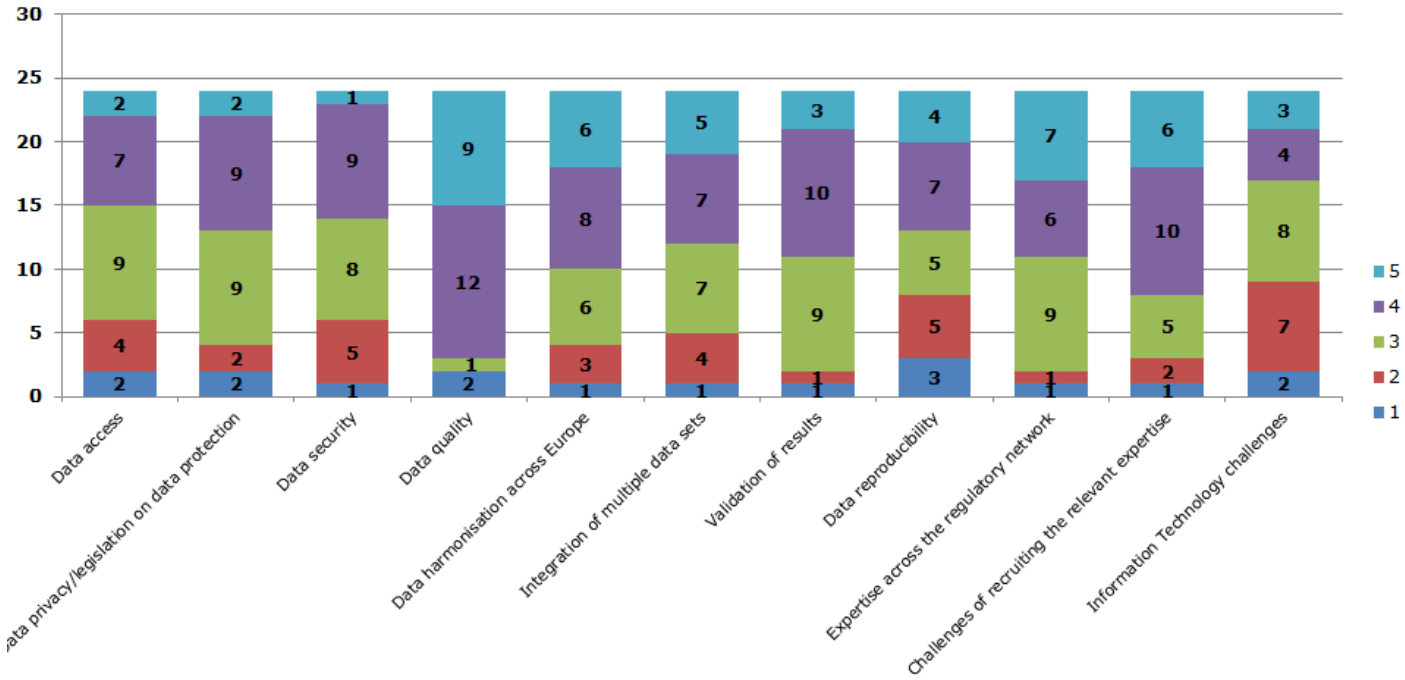


Mainly Signal detection

"Other areas":

- **RO:** IN HTA evaluation;
- **SI:** Pricing;
- **UK:** Principally post-licensing and to support development and monitoring of medical devices and in vitro diagnostic medical devices (IVDs);
- **IT:** Appropriate use and prescription of medicines in the real world evidence setting.

Q12 - What are the biggest challenges in the use of big data from the perspective of your NCA? Please prioritise your choices on a scale of 1 to 5 where 1 corresponds to the lowest challenge and 5 the biggest



4 + 5 > 2x(1 + 2 + 3) for:

Data quality

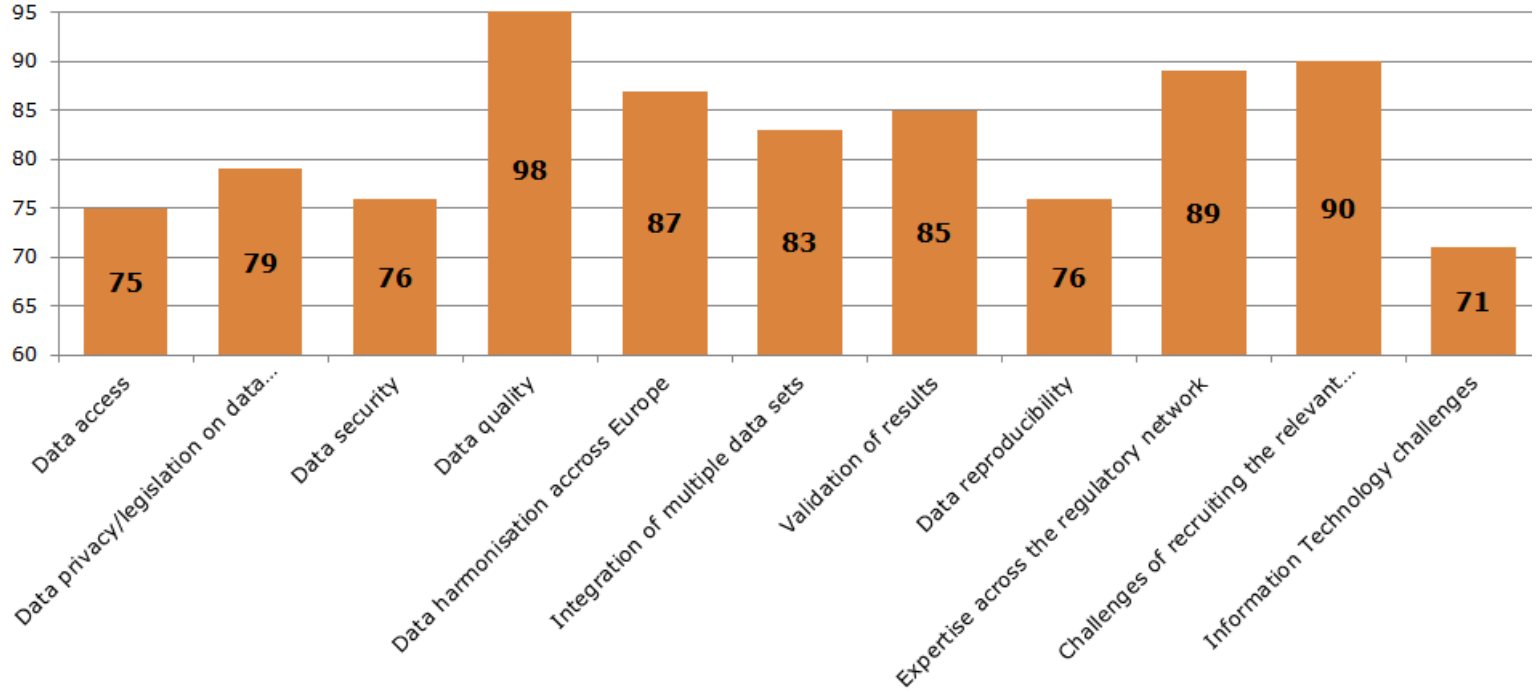
4 + 5 > 1 + 2 + 3 for:

- Data quality, Data harmonisation across Europe, Validation of results, Expertise across the regulatory network, Challenges of recruiting the relevant expertise

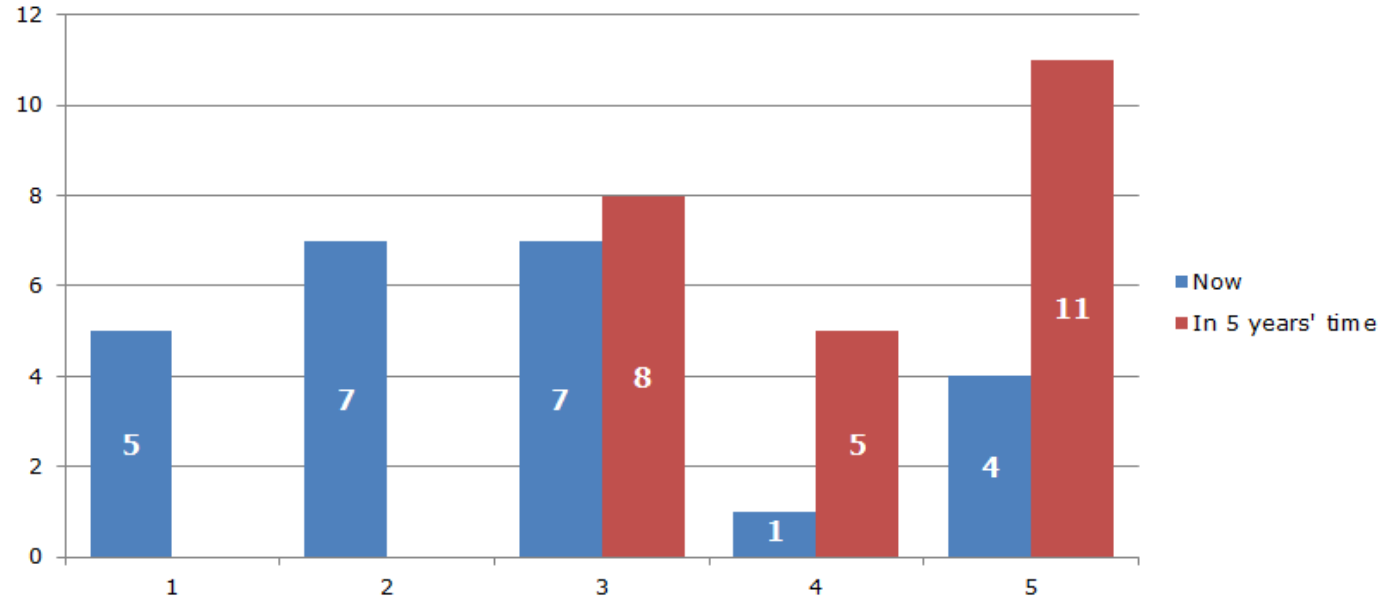
Comment from NL about data linkage: *genomic data is difficult to link to treatment data of CT*

Q12 - What are the biggest challenges in the use of big data from the perspective of your NCA? Please prioritise your choices on a scale of 1 to 5 where 1 corresponds to the lowest challenge and 5 the biggest

Scores from 1 to 5 multiplied by the number of NCAs for each area



Q13 - How important does your NCA management consider the use of big data to support regulatory decision making? Please indicate a number (1-5), where 5 is very important and 1 is not important.

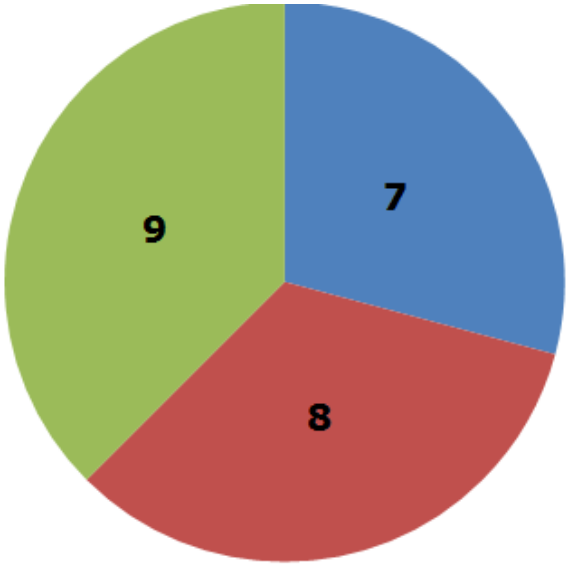


Main message: Use of big data to support regulatory decision making **is more important in the future**

Comments:

- **IS:** *This will depend on the development in the sector and number of cases including big data we will actually face at our agency.*
- **UK:** *Considerable variation in this depending upon the clinical area and the type of decision being made.*
- **DE BfArM:** *PV requires access to ADR databases*

Q14 - Is there any specific national legislation on access to big data in your country?



- Yes: Denmark, Finland, Norway, Austria, United Kingdom, Italy. Germany BfArM
- No: Malta, Czech Republic, Hungary, Spain, Netherlands, Slovenia, Belgium, Portugal
- Unknown: Germany BVL, Latvia, Germany PEI, Ireland, Slovak Republic, Iceland, Cyprus, Romania, Croatia

In which specific areas (6 NCAs):

- **DE, DK, NO, AT, UK, IT:** Data protection / privacy / access;
- **UK:** Recommendations for regulation of use of patient-identifiable information;
- **FI:** Biobanks and healthcare data;

Additional comments:

- **SI:** *We consider Big Data to be more relevant to larger NCAs. Big data processing is also dependent on national informatisation strategies and inter-organizational initiatives*
- **UK:** *Need to be clear on challenges/purpose before getting into details. Medical devices/IVDs are key in medicines development (stratification in clinical trials). Need to understand regulation of devices in generation/analysis of big data.*
- ***IT:*** *Currently Decre 196/2003, and the EU Personal Data Protection Regulation (Reg. 679/2016) as of 25 May 2018.*
- **DE BfArM:** *Access to healthcare data is covered by legislation on privacy data (Section 75 of SGB X; https://www.gesetze-im-internet.de/sgb_10/__75.html)*

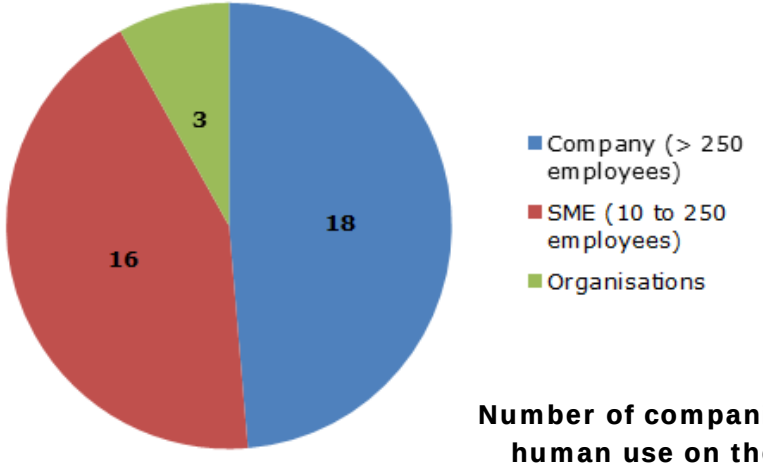
NCAs survey results: Conclusion

- **Currently very limited expertise** in big data analytics at national level;
- NCAs believe that **big data is not currently used** to support decision making on regulatory procedures;
- Although **more than half** of the NCA respondents have:
  - direct access to external big data sets (mainly RWD / Observational data, ADR data and CT data)
  - in-house systems / tools to analyse primary data sets (mostly SAS, SQL, R)
- Currently **little collaboration with academic institutions** (only 6/17 NCAs) and **outsourcing** of big data-related services (3/20 NCAs);
- **However**, the use of big data to support regulatory decision making is considered **important in 5y time**;
- And 2/3 of NCA respondents consider **an increase in FTEs** with expertise in big data necessary, despite **no concrete timescale set**;
- **Signal detection** is identified as the regulatory area where big data should be applied first;
- **Data quality** is the biggest challenge in the use of big data followed by availability and recruitment of expertise. **Data access/reproducibility** and Information Technology are the smallest;
- Mainly **unknown or no legislation** on access to big data.

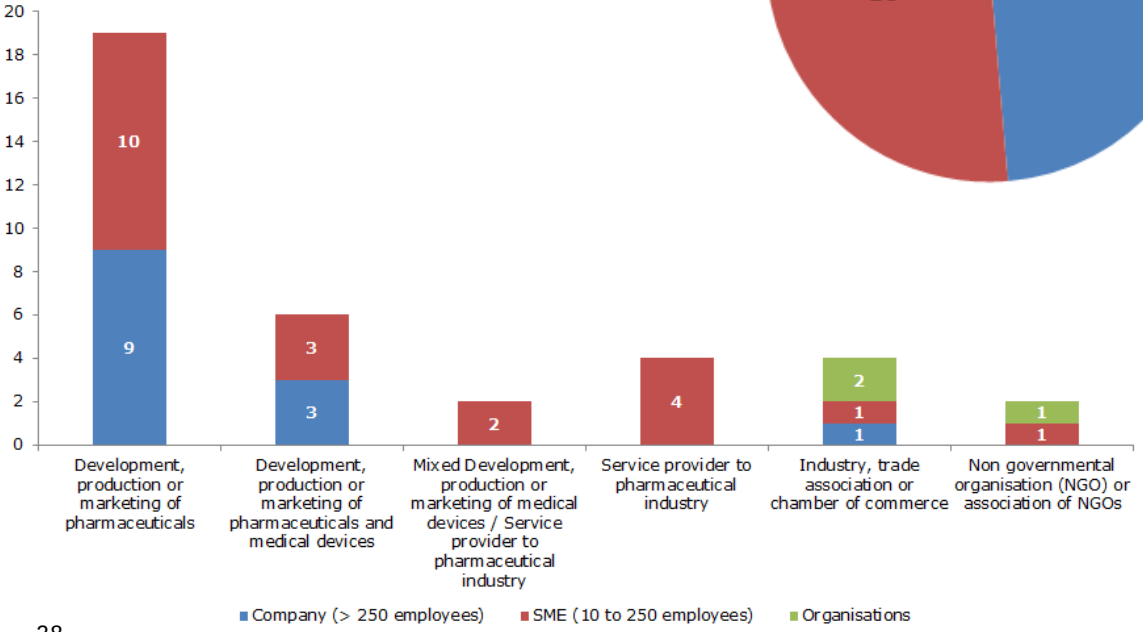
Pharma survey results

Q1 / 2 / 3: Profiles

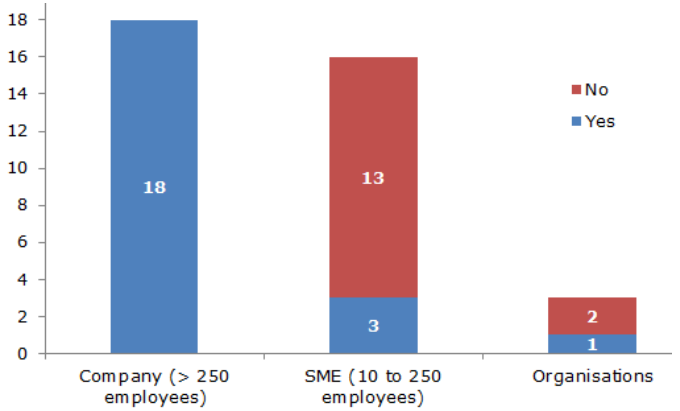
- Number of respondents: **37**



Number of companies per type of profile:

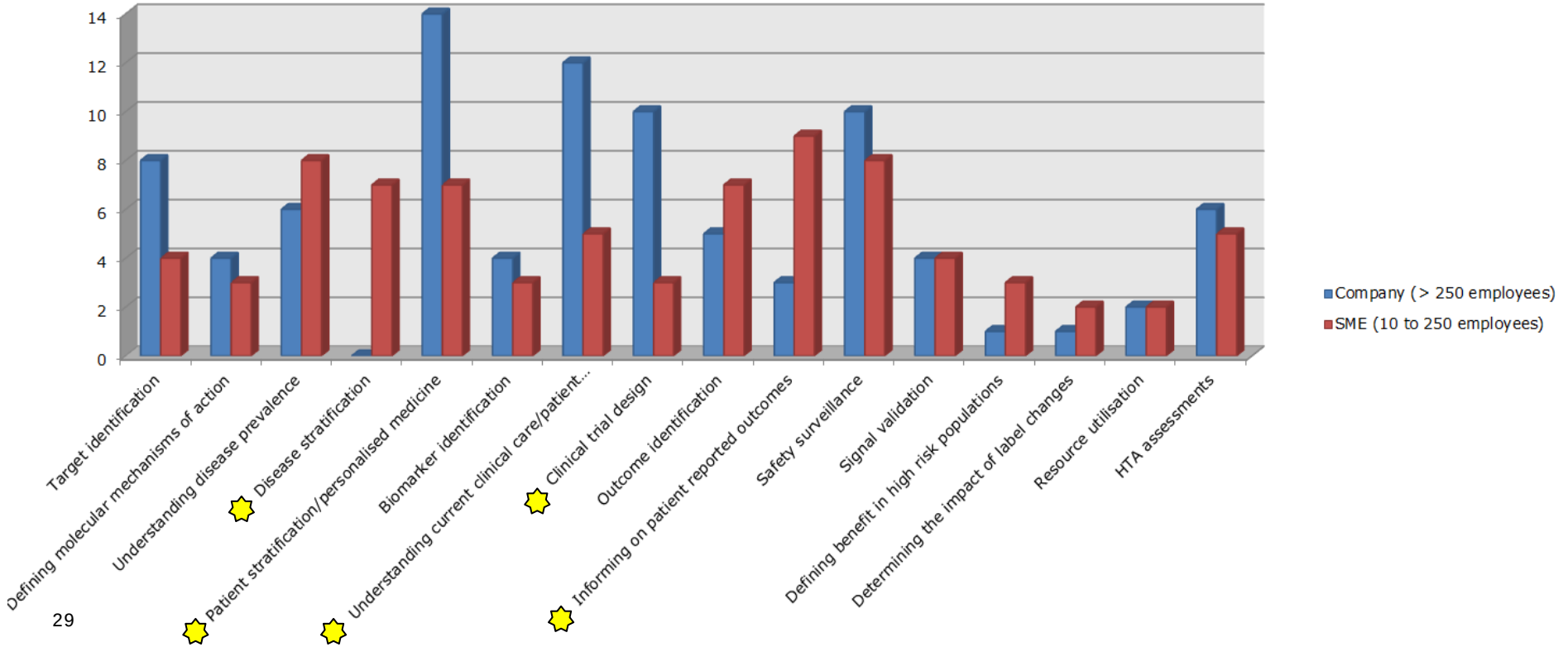


Number of companies with medicines for human use on the market within EU:



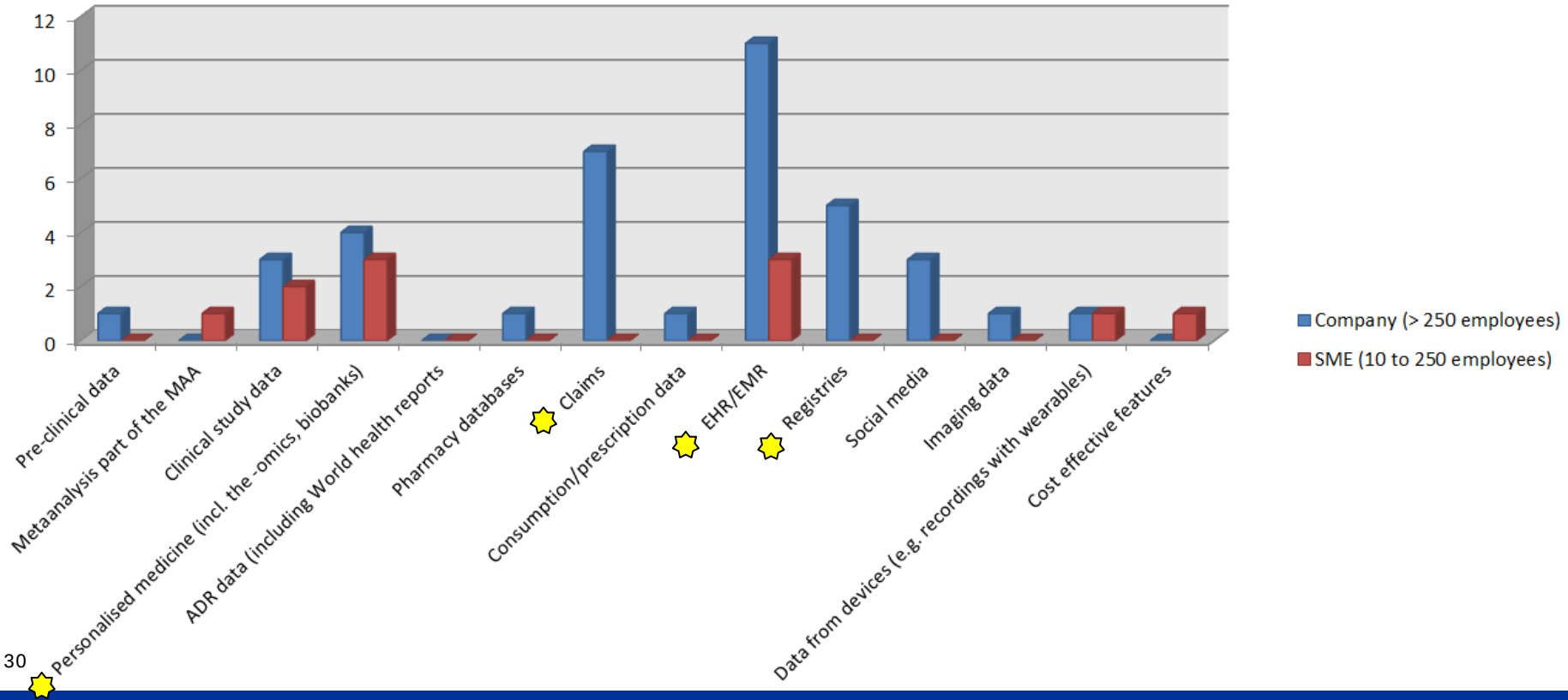
Q4: Key areas where big data will have the greatest impact

Number of companies that identified the following areas as being most impacted by the use of big data:



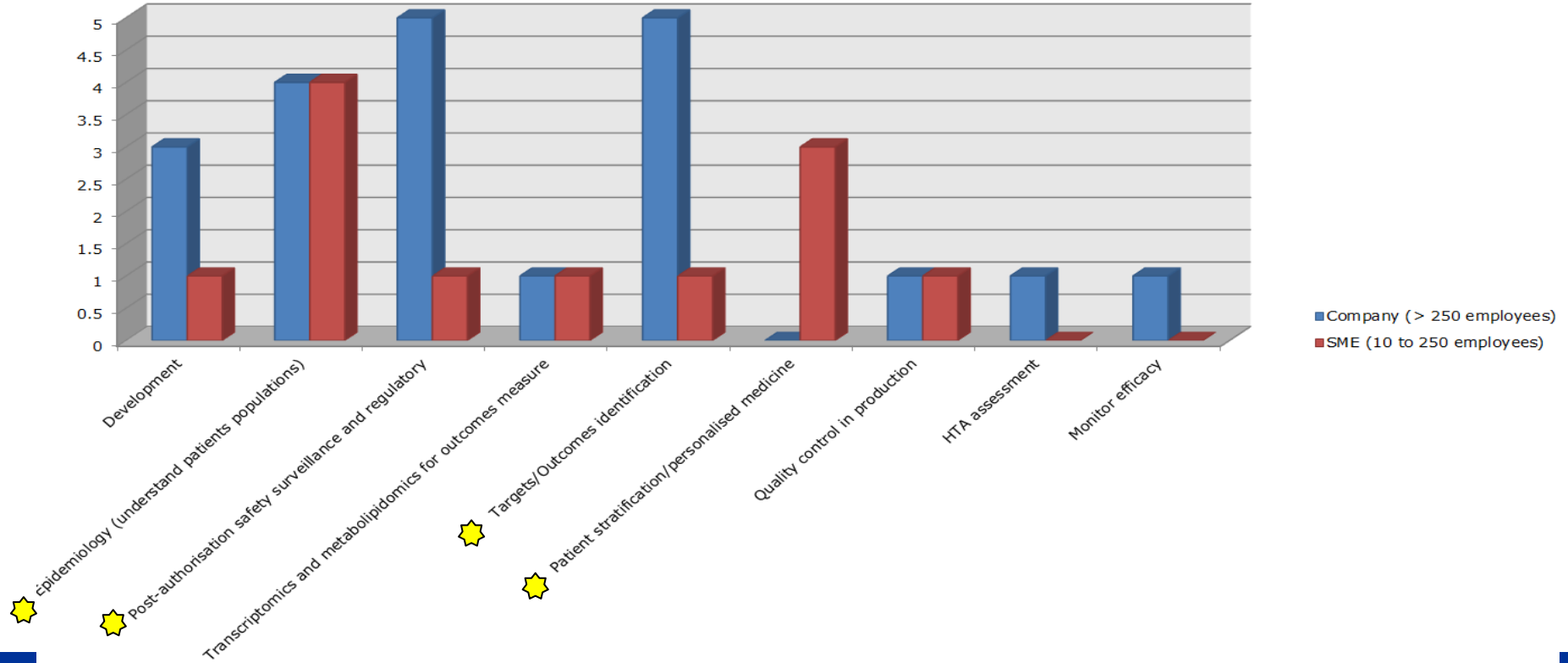
Q4: Datasets applicable to the impacted areas

Number of companies that identified the following datasets:



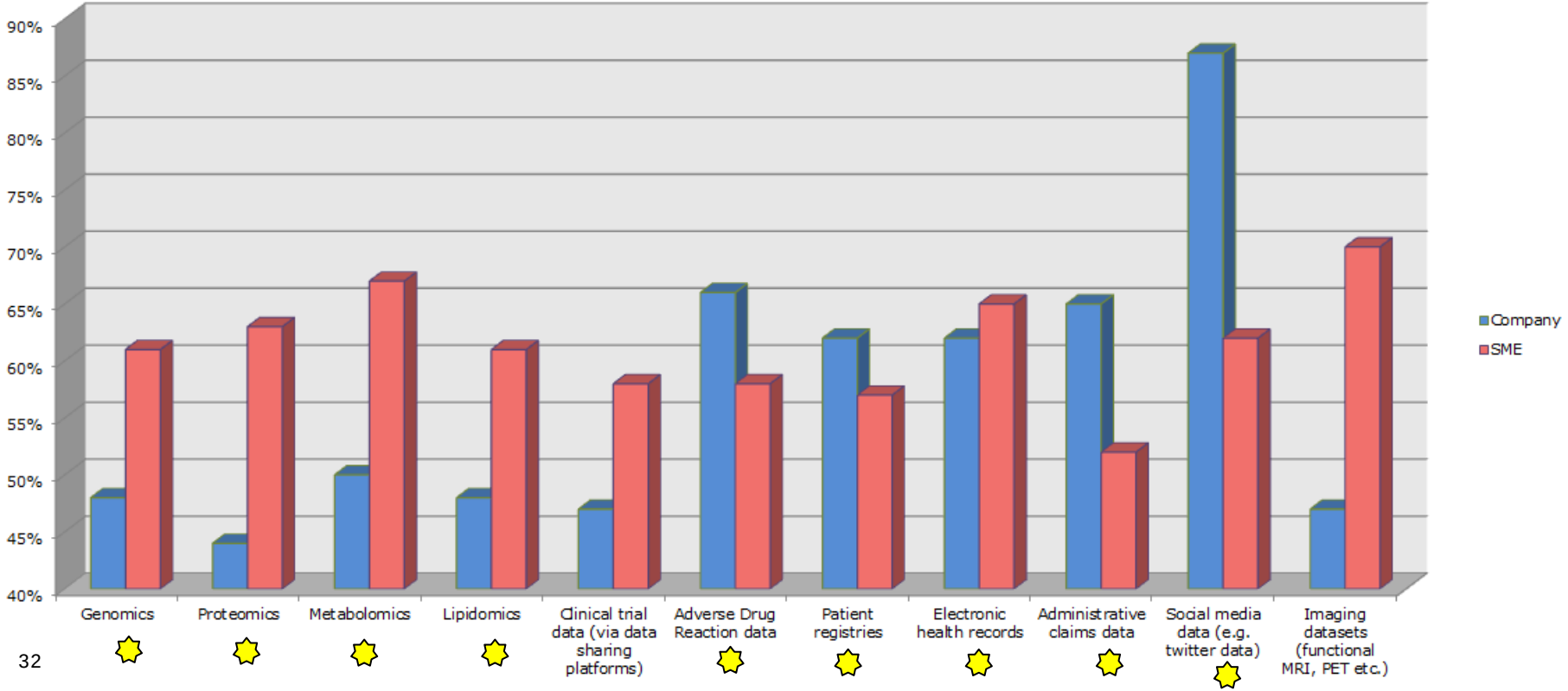
Q5: Examples where big data has been used to drive decision making across products' lifecycle:

Number of companies that identified the following areas where big data has been used to drive decision making:



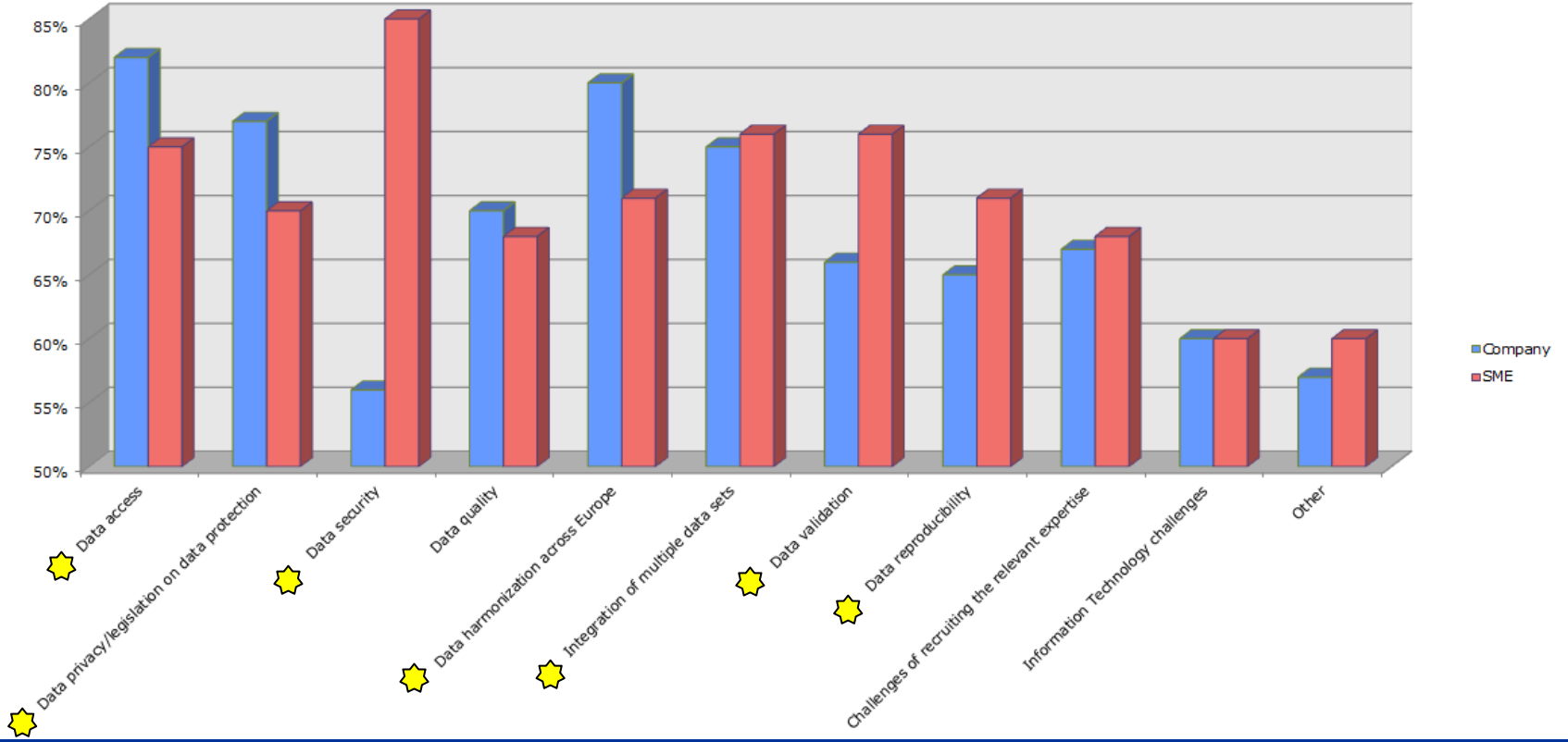
Q6: High level of concern on the validity of big datasets

Proportion of the cumulative highest score reached (where the highest score is 5x18 = **90** for Company and 5x16 = **80** for SME):

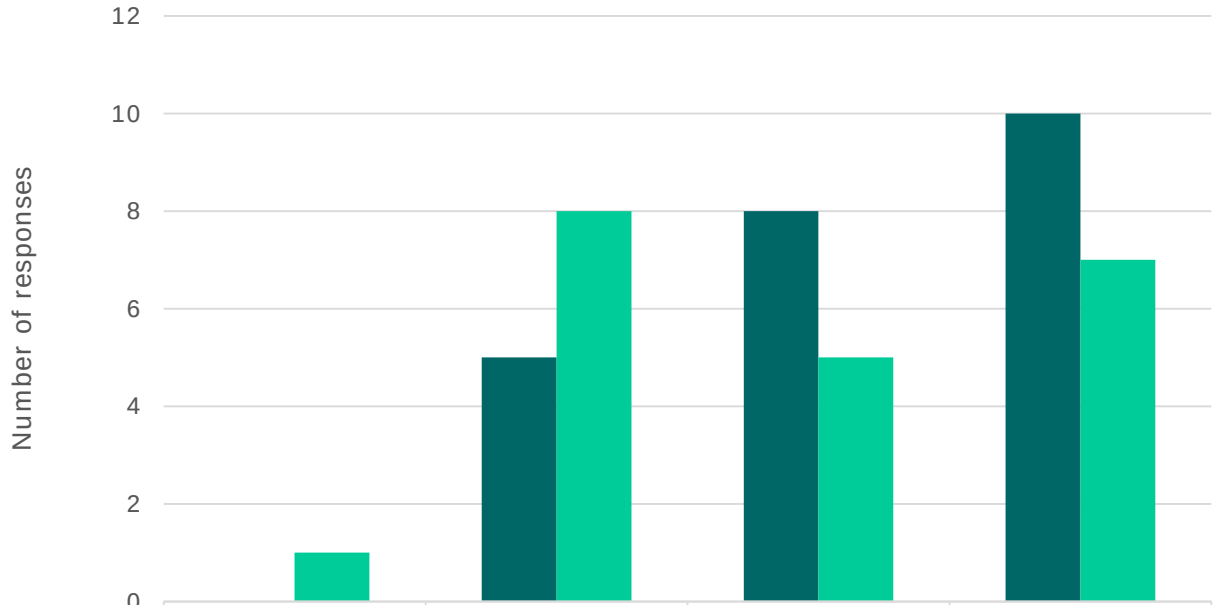


Q7: Key challenges in the use of big data

Proportion of the cumulative highest score reached (where the highest score is 5x18 = **90** for Company and 5x16 = **80** for SME):



Q8: What measures could the regulatory network introduce to address these challenges?

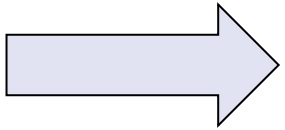


	Openness	Rules/Regulation/ Guidelines	Provide information	Provide data sources
Company	0	5	8	10
SEM	1	8	5	7

Q8: What measures could the regulatory network introduce to address these challenges?

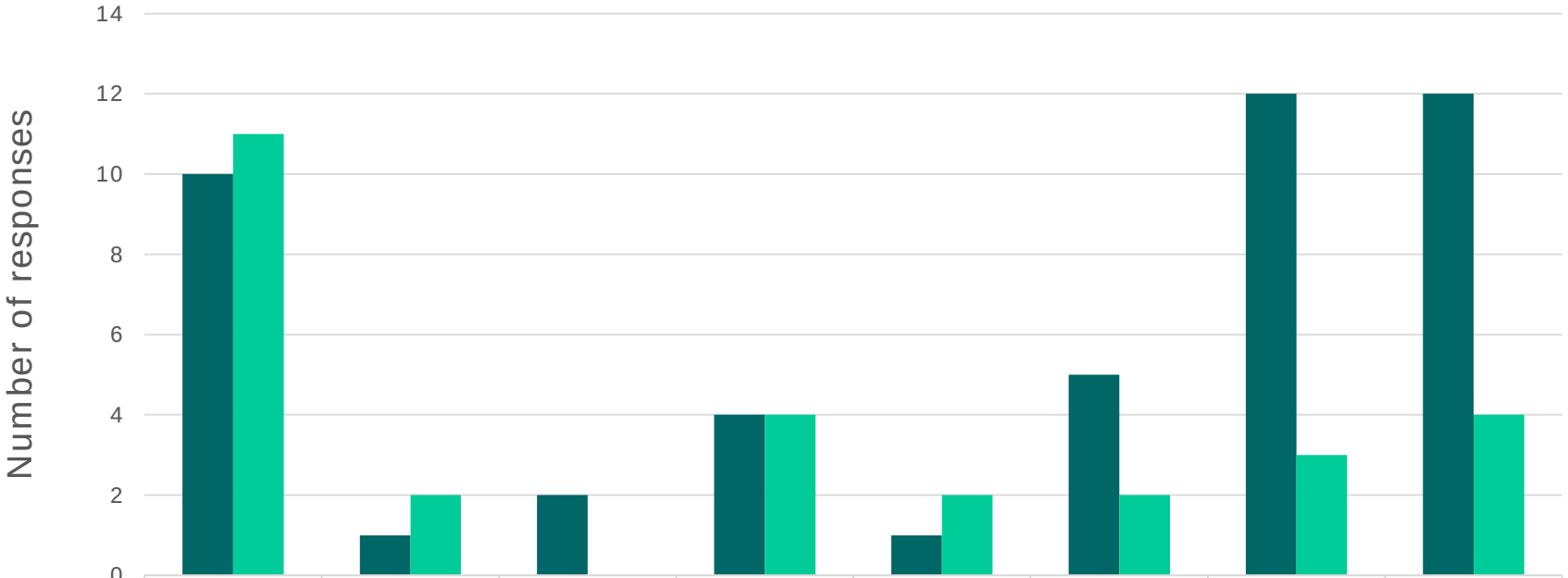
For rules/regulations/guidelines the following topics were mentioned:

- Data security / protection / privacy
- Harmonisation
- Data integration
- Standardisation and validation of endpoints
- Introducing measures to foster datasharing
- Defining data standards and quality requirements
- Guidance on usability of big data in regulatory decision making



Thereby, facilitating linkage

Q9: What are the greatest international challenges?



	Harmonization within and between countries	Cost reduction	Optimization process / speed up regulation	Data protection	Ownership	Big data analysis	Data quality	Access to data
Company	10	1	2	4	1	5	12	12
SEM	11	2	0	4	2	2	3	4

Q10: Any further comments

General comments on the survey and the definition of ”big data”:

- *The format of this survey does not allow for full discussion of the complex opportunities and challenges associated with the use of big data. In case of future opportunities, request for more free text input.*
- *Very broad definition of “Big Data” used in Survey. However, all these data sets are not ‘equal’ in terms of their regulatory acceptance. Further definitions would be appreciate.*
- *The HMA/EMA Task Force is addressing real world data; calling it a task force on “big data” is a bit of a misnomer.*

Q10: Any further comments

More specific comments:

- *Collaboration across stakeholders from regulatory, payer, providers, patients as well as industry will be key.*
- *The increased collaboration between all stakeholders is welcomed and should be continued.*
- *Openness to new ideas*
- *Data anonymisation, harmonisation standards, as well as trusted technologies to ensure data privacy and protection are needed in the EU: “It will be a fine balance in keeping the personal ´data protected and at the same time make it available.”*
- *EMA should get more actively meaningful patient centered outcomes - based on mobile sensors. Focus on a new concept like "mPRO”, i.e. "mobile device based patient reported outcome measures".*

Pharma survey results: Conclusion

	Companies (> 250 employees)	SMEs (< 250 employees)
Greatest impact of big data:	• Target identification • Patient stratification and personalised medicines • Post-authorisation safety	• Outcome identification • Informing on patients reported outcomes • Diseases prevalence
Highest concerns on the validity of big datasets:	• RWE data sets • Social media	• "-omics" • Imaging datasets
Key challenges in the use of big datasets:	• Data access • Data privacy • Data harmonisation	• Data security • Data validation • Data reproducibility
Greatest international challenges:	• Harmonisation on many aspects within and between countries including on access rules, data protection/privacy, data standards, collection, validation.... • Data quality • Data access	
Regulatory measures to address these challenges:	• Need for clear regulatory guidance (including on usability of big data in regulatory decision) for better harmonisation (see above row) • Facilitation of access to the data, fostering data sharing	

Thank you, any questions?

Further information

[Insert relevant information sources or contact details as applicable.]

European Medicines Agency
30 Churchill Place • Canary Wharf • London E14 5EU • United Kingdom
Telephone +44 (0)20 3660 6000 **Facsimile** +44 (0)20 3660 5555
Send a question via our website www.ema.europa.eu/contact

Follow us on **@EMA_News**