Making the case for the collection of a minimal dataset for children with speech sound disorder

Abstract

Background: NHS case note data is a potential source of practice-based evidence which could be used to investigate effectiveness of different interventions for individuals with a range of speech, language and communication needs. Consistency in pre- and post-intervention data as well as collection of relevant variables would need to be demonstrated as a precursor to adopting this approach in future investigations of SLT intervention.

Aims: The aim of this paper is to explore whether routine clinical data collection for children with speech sound disorder (SSD) could be a potential source for examining the effectiveness of intervention(s).

Methods and Procedures: We examined case notes from three UK NHS services, reviewing 174 sets of case notes and 234 blocks of therapy provided for school-age children with SSD.

Main contribution: We found there was significant variation in pre- and post-intervention data and variables collected by the services. The assessment data available in the case notes across all sites were insufficient to be used to compare the effectiveness of different interventions. Specific issues included lack of consistent reporting of pre- and post-intervention data, and use of a variety of both formal and informal assessment tools.

Conclusions and implications: The case notes reviewed were from three sites and may not represent wider clinical practice nevertheless the findings suggest, the sample explored indicates the need for more consistent and contemporaneous collection of data for children with SSD, to facilitate investigation of different interventions in practice. Researchers should work with the clinical community to determine a minimal dataset that includes a core outcome set and potential variables. This should be feasible to collect in clinical practice and provide a dataset for future investigations of clinically relevant research questions. This would provide an invaluable resource to the clinical academic and research communities enabling research questions to be addressed that have the potential to lead to improved outcomes and more cost-effective services.

What this paper adds

What is already known on the subject

While there is some evidence for the efficacy of therapy for children with SSD, studies typically focus on very specific populations who meet strict selection criteria and take place in university clinics or laboratory style settings which do not reflect typical clinical practice in the UK and elsewhere. An alternative approach to investigating the effectiveness of interventions would be to use NHS case note data. It is not clear from the existing literature whether case note data are sufficiently robust to facilitate such an analysis.

What this paper adds to existing knowledge

This study found that case note data, in particular assessment data, were highly variable across services and would be insufficient to compare different interventions for this population. Agreement on what should be included in a minimal dataset for children with SSD is required to maximise the potential for NHS clinical case notes to become a resource for future research.

What are the actual or potential clinical implications of this work?

This study indicates that current clinical practice in SLT for children with SSD is inconsistent with regards to the reporting of pre- and post-intervention assessment data and other important variables in case notes. We make the case for agreeing a minimal dataset, with a need for clinicians to work with researchers to determine core outcomes and additional relevant data, which can be feasibly collected in clinical practice.

Introduction

Evidence for effectiveness of intervention is a cornerstone of speech and language therapy (SLT). We need this to inform our practice and to convince commissioners of the value of our services but determining what information to collect is challenging. Hoffman et al. (2010) present a model of evidence-based practice which describes using clinical reasoning to integrate information from four sources: research evidence, clinical expertise, the patient's values and circumstances, and the practice context. While the combination of these sources can be useful in guiding clinical care, it is nevertheless the best research evidence which provides the most robust data to inform practice and service design. Research evidence, as reported in the academic literature for SLT, varies across clinical groups but in many cases, the strength of the evidence is low with few interventions having been studied sufficiently to be explored through meta-analyses. A number of systematic reviews have found this to be the case for children with speech sound disorder (SSD). For example a review of treatment intensity by Kaipa and Peterson (2016) found only one controlled trial. Another review considering the involvement of parents in intervention for SSD noted the need for high-quality research which reports more detail regarding the involvement of parents and home tasks in intervention (Sugden et al. 2016). A systematic review of interventions for speech-sound disorder in preschool children (Wren et al. 2018b) found case series was the most common research design. Reviews of single intervention approaches have fared little better, in a review of ultrasound biofeedback the most common research design that met the inclusion criteria was single-case experimental design, (Sugden et al. 2019). Another intervention, non-speech oral motor exercises, has been subject to three systematic reviews and all reached the same conclusion - the evidence available was insufficient to either recommend or discourage use of this approach (Lee and Gibbon 2015, McCauley et al. 2009, Ruscello 2010).

While a hierarchy or pyramid of evidence, such as that presented by Greenhalgh (1997), is the traditional way that levels of evidence are evaluated in intervention research, an alternative approach is suggested through Fey and Finestacks' (2009) 'phases of intervention research'. The five-phase model for studying the effectiveness of intervention begins with the "pre-trial" studies in Phase I. In Phase II "feasibility studies", the clinical viability of the intervention is tested. Case studies, discovery-oriented single-subject designs, and small group cohort control studies are all appropriate at this stage. Phase III "early efficacy" studies intend to test efficacy. Many types of designs are appropriate at this stage but Fey and Finestack

(2009) note they should be experimental, or at least quasi-experimental. Phase IV research or "later efficacy" aims to directly compare the target intervention with an alternative intervention or a no-treatment condition to address causality but under more generalizable conditions. Finally, "effectiveness" research, is determining whether "the therapeutic effect is realized in day-to-day clinical practice" (Robey 2004, p. 405). Generally, Phase V research is considered "field research" or "community-based research".

Randomised controlled trials (typically phase IV of Fey and Finestacks' model) are generally considered the gold standard for intervention studies and have been used in a number of high-profile studies over many years in SLT (Bowen et al. 2012, Murray et al. 2015, Palmer et al. 2019, Glogowska et al. 2000). However, questions about how and when trials for SLT should be used have been raised (Pring, 2004). Problems have arisen when randomized control trials are used to examine therapy provision for a specific client group, since clients are often a heterogeneous group and who require individualised, and hence different, intervention programs (Baker and Mcleod 2004). Indeed, a feature of speech and language therapy is that interventions tend to be complex; utilising multi-method approaches or programmes, making the assessment of efficacy of one approach difficult.

Whilst well designed and adequately powered randomised controlled trials provide the most robust support for an intervention, there are other approaches which can be used to help build the evidence beyond single case and small group studies. Observational cohort studies are situated below trials in the levels of evidence pyramid (Dobinson and Wren 2019, Murad et al. 2016), but when data are collected prospectively, they can provide a large dataset to explore a range of potential research questions, including those regarding the effectiveness of specific interventions. Observational studies also have the benefit of being more reflective of every day clinical practice, as demonstrated by Taps (2008) in a study of provision in schools. While Fey and Finestack (2009) situate studies of effectiveness in practice as the final phase of research, there are times where such observational designs are more feasible and appropriate than RCTs.

Practice Based Evidence (PBE) research methodology is an observational approach that can happen alongside usual practice and can be seen as the systematic collection of measures associated with particular treatment goals or desired outcomes, with the potential to examine associations among many treatments and outcomes. This approach has already been used to inform and tailor everyday clinical practice in the context of client characteristics in other fields, for example in spinal cord injury (SCI) rehabilitation (Whitneck and Gassaway 2013). 1500 patients with SCI were recruited and the rehabilitation process was captured by clinicians from

multiple disciplines. Outcome data were abstracted from medical records and obtained from patient interviews at 6 and 12 months. The application of PBE methodology in the SCI rehabilitation study provided extensive information about the process of inpatient SCI rehabilitation. From these data, researchers were able to identify factors that impacted outcomes from both demographic and injury characteristics type as well as quantity of treatment interventions.

In the UK, we have the environment to support a natural cohort study through the NHS where we have the potential to collect routine data in clinical practice as part of establishing PBE. Indeed, SLT services for children born with cleft palate are doing this already. Protocols for routine collection of data have been agreed nationally and are being used to measure outcomes against national standards, compare outcomes across services and address other clinically driven research questions (Wren et al. 2018a).The cleft population is relatively small and more easily defined than the SSD population, which arguably makes the process of data collection more straightforward for cleft services than might be the case for other groups. Nevertheless, the routine data collection within cleft services provides a precedent that a similar approach could be applied to other clinical groups in SLT such as SSD, with the possible gain of a large dataset which can be mined to investigate the impact of different interventions on outcomes. Consideration of variables such as the nature and severity of the SSD, age, gender, SES, and co-morbidities could provide further valuable information to investigate in analysis. As electronic health records become more common place across the world such approaches should also become more readily accessible and achievable.

The challenge for any such undertaking will be the variation which currently exists in how services are delivered, and even more importantly, in what, when and how data are collected. The Royal College of Speech and Language Therapists' (RCSLT) work on establishing a tool for outcome measurement across the profession has passed a milestone in obtaining agreement for a national outcomes tool which has been successfully piloted (Moyse et al. 2020, RCSLT Online Outcomes Tool (ROOT 2019). The ROOT is based on Enderby and John's (2015) Therapy Outcome Measures (TOMs) and can be used to collect data identifying outcomes across client groups and services. It was intentional that this tool should be universal to maximise take up and enable as many services as possible to engage with the activity. However, TOMs is less suitable for addressing research questions around variation in response to intervention for specific clinical groups because of its lack of sensitivity for disorder-specific outcomes. Core Outcome Sets (COS) have been developed for some clinical groups such as

aphasia (Wallace et al. 2019), but the primary purpose of these is for use in trials of intervention which sit outside of usual clinical practice. As a consequence, they do not typically translate well to the clinical setting where practice and time spent on intervention may significantly vary.

Using ROOT to measure outcomes across a service is ideal as it allows for comparison and identification of which types of provision are leading to better and more cost-effective outcomes. However, more detail is needed to address questions relating to intervention for disorder-specific groups such as SSD. Outcome data such as that from the ROOT can be limiting, as they are unable to take account of all variables, particularly the nature of impairment, and in the instance of SSD, the different sub-types and error patterns for a particular child. Other variables such as socio-economic status, gender and age as a minimum but also possible co-morbidities also need to be considered depending on the clinical group of interest. Collection of a minimum dataset containing core outcome data and a set of common variables that impact practice is needed in order for the resource to be used to address questions which have clinical relevance and impact.

Children with speech sound disorder

This population shares many features with children born with cleft palate and is therefore an ideal group to explore whether routine case note data could be used as a data source for clinical research questions regarding effectiveness of interventions. A wide range of interventions have been investigated already for the SSD population and are reported in the literature. However, intervention trials reported to date tend to involve participants receiving significantly more therapy than is offered in everyday practice (Almost and Rosenbaum 1998, Gillon 2000). Indeed, in a recent systematic review Wren et al. (2018b) found that while studies provided a wide range of number of sessions (3-67), over 60% of the studies provided therapy either two or three times a week. Moreover, the duration of therapy varied from six weeks to six months, with most providing therapy for more than eight weeks. This contrasts with findings from a recent UK survey of current SLT practice for children with SSD where the most common frequency of intervention was once weekly delivered mostly across a range of 9-12 sessions (Hegarty et al. 2018). Indeed, as McCartney et al. (2011) indicate, the total number of sessions provided within NHS community services is often capped at 6 weeks.

In order to determine whether intervention for children with SSD is effective in typical clinical settings with typical resource allocation therefore, there is a need to use an alternative approach. Given the potential described above for the NHS to function as a natural cohort study,by creating a resource of practice-based evidence, it may be possible to use case note data collected in routine clinical practice. If this approach is feasible, case notes could be a viable source for investigations within the NHS with the potential for large datasets and adequately powered samples. The first step towards this is determining the extent to which services are consistent in following clearly defined and robust procedures with respect to data elicitation and case note documentation.

Aim

The aim of this paper is to explore whether routine clinical data collection for children with speech sound disorder (SSD) could be a potential source for examining the effectiveness of intervention(s).

Method

We used retrospective data from an evaluation of case notes from three NHS services in the South and South West of England. The three sites covered mixed urban and rural populations. Informal conversations also took place with service managers and some SLT staff about the nature and set up of their services.

Data collection

The inclusion criteria for children whose notes were included in the study were that they were aged between 4 years, 0 months and 8 years, 11 months old and presented with moderate-severe difficulty with speech on assessment, based on clinical judgement of the therapist carrying out the assessment. Children with known neurological deficits, sensorineural hearing loss, structural anomalies or pervasive developmental disorders were excluded, although those with language difficulties were not.

Data on each participant's pre- and post-intervention assessment measures were collected together with information on the type of provision received for blocks of therapy. Blocks of therapy refer to a discrete period of therapy, usually starting with an assessment, with therapy delivered regularly during the block (usually for a period of about 6 weeks). At site A, only data on the most recent block of therapy that targeted speech were available whereas at sites B and

C, data from all blocks of therapy that targeted speech were collected. At each site data were collected in pairs, this was mostly completed by researchers although at site A and B clinicians assisted. If the clinical notes referred to assessments but copies of assessment data could not be located they were reported as not completed. All identifiable information was redacted from photocopied assessments and other relevant data were anonymised and taken for all participants for reliability checking. Table 1 summarises the data that were collected from each site.

[Table 1 about here]

For each set of clinical records, post intervention data were collected in whatever form was available (e.g. published screening tools as well transcribed samples of the child's speech from informal assessment pre- and post-intervention). A record of the assessments used pre- and post-intervention for each block of therapy was made as well as any additional speech assessments that were carried out. Data were also collected on the therapy received and method of service delivery including when it was delivered, who acted as the agent of therapy, and the duration of the therapy block in terms of the number and frequency of sessions.

Main Contribution

The three sites used different combinations of service delivery methods. Table 2 provides a summary of these data, together with percentages for each pattern of delivery for each site. Sessions delivered by SLTs were typically delivered weekly or every other week, with number of sessions varying widely (typically 2-12). Children in language units received up to 33 sessions. Case notes of children who had received blocks of therapy from teaching assistants in schools provided very little documentation about the intervention received, except information on the sounds targeted, and no information about the number, frequency or duration of sessions that the child received.

[Table2 about here]

Availability of data

Data available pre- and post-intervention

There tended to be more transcribed data available pre-intervention than post-intervention. Transcribed data included phonetic transcription of any single words (and connected speech data if available), where the target word (s) were evident. This varied by site, however. Table 3 summarises the transcribed assessment data that were available across sites (note these were mostly single word data with lists covering most consonants and word positions). For all sites, fewer children received post therapy assessments, although this difference was greatest for Site C.

[Insert Table 3 here]

Data available by proximity to therapy block and service delivery method

Figure 1 summarises the transcribed assessment data obtained from the three sites with regards to the different service delivery methods for each of the blocks of therapy and the availability of pre- and post-intervention assessment data within both one and three months of the first and the last intervention session. Out of a total of 234 blocks of therapy, pre- and post-intervention assessment data within one month of the therapy block were only available for 33 blocks (less than 15%). While slightly more were available within three months (N=60, 26%), data from case notes were still missing for most of the sample (N=173, 73.9%).

[Figure 1 about here]

Types of pre- and post-intervention data

A total of ten different assessments were used across the three sites. All of these collected samples of single words (see Table 4). One was a standardised assessment (Diagnostic Evaluation of Articulation and Phonology (DEAP) (Dodd et al. 2002); four were published but not standardised; CLEAR Phonology Screening Assessment (Kerryjane and Spilsby 2006), South Tyneside Assessment of Phonology (STAP/STAP2) (Armstrong and Ainley 1992; 2012), Phonology Screening Assessment (PSA) (Stevens and Isles 2001) and The Nuffield (Williams and Stephens 2004). Another five assessments were categorised as informal picture naming assessments, which were developed in house and hence specific to individual services. An exception to this was the Bowen quick screener (Bowen 1996).

All of these assessments elicited single word data which were transcribed using broad phonetic transcription. Although some of these assessments come with analysis sheets, these were only completed for two children across the whole sample and there were no further records of any analysis in the notes. The Phonetic and Phonological Systems Analysis (PPSA) (Bates and Watson 2012) was also completed for two children. For 15 (6%) episodes of care there was explicit recording of phonological processes or comments about processes. Findings indicate that clinicians were using assessments primarily to elicit a speech sample which they then inspected visually to identify common phonological processes and phonetic inventories or to determine a child's knowledge of phoneme classes. While the exact way these were used isn't known it is evident that the children's targets were determined by their performance on the assessments.

[Insert Table 4 here]

At Site A, the CLEAR Phonology Screening Assessment was the most commonly used pre-(n= 42, 34%) and post-intervention assessment (n=32, 26%), with clinicians providing phonetically transcribed responses (not a requirement of this assessment). At the end of therapy blocks, targets set at the beginning of intervention were reviewed as 'achieved', 'partially achieved' or 'not achieved' based on either formal assessment or clinical impression.

At Site B, at pre-intervention assessment the majority of children were assessed with the PSA (n=18, 56%). Post-intervention, progress for the majority of children was determined by the degree to which targets set at the beginning of the intervention block had been achieved (n=18, 56%) rather than through comparison of transcribed speech samples from single word or connected speech assessments. For example, a target such as 'for Jon to produce /l/ at the beginning of five different words'.

At Site C, four different developed in house (none published) assessments were used. These all included assessment of single word production with accompanying phonetic transcription. Pre-intervention the majority of children (n=52, 63%) were assessed with in house assessment we have named "Assessment A". Although children were frequently seen for review post-intervention, assessments which included phonetic transcription were only conducted for half of the children (n=41, 50%). Assessment A remained the most popular post intervention assessment (n=20, 25%).

Summary

The aim of this paper was to explore the content of case note data for children with SSD to determine its potential for examining the effectiveness of intervention. We examined 174 sets of case notes which reported on 234 blocks of therapy. We found that pre- and post-intervention assessment data (collected within one month of a therapy block) were only available for 33 blocks of therapy across the three sites. There was variation in whether the same pre- and post-intervention assessments were conducted, and also in the degree to which transcription was carried out. Ten different assessments were used across the sites, one of which was the only standardised assessment in the UK for speech in children aged 3;0-6;11 (DEAP, Dodd et al. 2012).

Evaluation of speech and language therapy intervention for children with SSD remains an important area for research. This is especially true in a climate where service leads rely on demonstration of effectiveness of SLT to fund services. However, this report based on three NHS clinical SLT sites demonstrates that data gathered from case notes is not currently sufficiently detailed or consistent enough to allow a comparison of the effectiveness of different interventions.

Data available in clinical SLT records

At the three sites explored here, a range of assessments were used in practice, including informal and in house assessments. There was also variation in terms of whether the same pre- and post-intervention assessments were conducted, and if transcription data were collected at all. It is noteworthy that while there was variability within services, greater difference was observed between services. Each site had a different assessment that was found to be used most frequently; no sites recorded STAP (Armstrong and Ainley 1992; 2012) as a frequently used assessment, an assessment that was previously reported to be most popular in a survey of SLT practice (Joffe and Pring, 2008). At Site C, the use of informal assessments developed in-house indicates organisations are taking a flexible approach to collection of assessment data. While these informal assessments may still be able to yield representative samples there was no recorded indication that they were being evaluated in a systematic way.

An intrinsic aspect of speech assessment is the need to transcribe what is heard and analyse the sample for errors and patterns. The Child Speech Disorder Research Network (CSDRN) has developed two sets of guidelines to support speech and language therapists in their transcription and analysis of children's speech samples (Bates et al. 2017; CSDRN, 2017).

However, observation of the case notes for this study suggests that these guidelines have not been implemented by speech and language therapists in clinical practice. It may be that clinicians are unaware of these guidelines, or feel too constrained by time pressures to follow them. Or it may be that because they have not been adopted at a national level they are not embedded in service systems which support clinicians to conduct best practice. This is in contrast to the national protocols and reporting mechanisms within speech and language therapy for children with cleft palate which have enabled smooth integration of minimal data sets and best practice for this population (Wren et al., 2018a).

Sufficiency of data to facilitate comparison of different patterns of service delivery

While the variety of assessments used in clinical practice makes evaluation of intervention challenging, the lack of consistent reporting of pre- and post-intervention data in the case notes we reviewed is especially limiting. This was true across all three sites explored in the present study, suggesting that current practice in routine data collection is insufficient for SLTS to evaluate whether therapy provided is effective for groups of children, but arguably for individuals as well. This lack of consistency also makes it difficult to know how SLTs are making differential diagnoses, prioritising therapy targets or indeed selecting an appropriate intervention approach. Assessment data are particularly pertinent when considering the impact of intervention on generalisation across the speech sound system and into communicative contexts, which is one of the key goals of phonological/pattern-based intervention. Without these assessments it's also unclear how children are meeting the conditions for discharge.

There could be a number of reasons for variability in what assessment data are collected in practice. At all three services there were more documented pre-intervention assessments than post-intervention assessments. While some post-intervention data were widely available, these were often reported in the form of 'targets met' rather than following a reassessment of a pre-intervention measure. The use of informal assessments such as observations of whether a child has met their target(s) is likely to have practical utility in the clinical context but provides limited objective evidence to determine whether the intervention provided has been effective. The use of in-house and informal picture naming tests may also be driven by what is available to practitioners. Funding to purchase assessments is limited and some clinicians and teams have responded by developing their own resources. Non-standardised and informal picture naming tests may not be problematic provided they result in a representative transcribed sample which can be used in analysis rather than a checklist or 'targets met' response. It

should also be noted that even if clinicians were only using published assessments this would not necessarily provide sufficient data for robust evaluation, for example assessments such as the CLEAR do not require transcription, and others, such as the STAP do not appear to provide a representative sample.

The reasons for a gap in the proximity of assessment to blocks of therapy appeared to be partially process driven. For example, conversations with SLT services indicated children are often referred, assessed and then placed on a waiting list for therapy, thus creating a gap between the initial assessment and the start of intervention. Moreover, anecdotal report from services indicated they felt it was important to spend time in the final intervention session on therapy rather than on an assessment when the next block of therapy may not occur for some time. Some children were also intermittently reviewed and assessed for general progress and appropriateness of targets, but not necessarily with the aim of establishing the effectiveness of a block of therapy per se.

Collectively, the review of case notes shows a reliance on tools that fall short of the extent and size of speech sample recommended (Bates et al. 2017; CSDRN, 2017) both to inform a robust differential diagnosis of subtype of SSD and subsequent selection of intervention approach and targets. Such informed clinical decision making underpins effective and efficient therapy provision. Whilst it was not the aim of this work to investigate this, it is a worrying finding nevertheless and possibly indicative of the likely resource constraints to services and high demand.

Looking forward: a minimal dataset

While there are clinical guidelines for assessment and transcription of speech, there is no national standard for data collection in NHS Speech and Language Therapy services. This shows in the lack of robust assessment data and consistency of data collection in the three NHS sites observed in this work. With no national standard for measurement of speech for children with SSD there is nothing for services to benchmark their assessment data against. Some variability between services in data collection practices is not necessarily problematic if they are collecting the essential data required for children with SSD, and data are being collected within certain time frames. Although we need to know how tools compare in terms of the sample they yield and how the data are interpreted, consistent collection of transcribed speech samples will permit analyses using tools such as the PPSA (Bates and Watson 2012) or measures within this such as Percentage Consonants Correct (PCC), even if the stimuli used to collect the samples

vary. In contrast, when outcomes are limited to a report that targets have been met, without supporting evidence in the form of recordings or transcripts of children's speech, there is insufficient evidence to conduct a robust analysis of the effectiveness of intervention.

It is noteworthy also that while measures such as PCC are useful they need to be considered in the contexts of pre- and post-intervention samples matched in terms of the phonemes' positions, distribution of consonants and phonetic contexts tested. For children with more severe SSD, it is likely there is a need for some qualitative data (i.e., natural phonological processes and atypical errors, contrastiveness and patterns of variability) as well as quantitative data (e.g. PCC). Meaningful comparisons using quantitative data alone may not be sufficient because not all errors are equal in terms of either the phonetic 'distance' between the target and its realisation, or the extent to which intelligibility is impacted. It is also difficult to know how results reflect changes as a function of targeted intervention without details regarding therapy targets or approach. Importantly, an agreed minimal data set would need to use a post-intervention measure directly comparable to targets assessed pre- intervention in order for outcomes such as PCC to have real meaning.

Exploration of three NHS sites indicate the need for a more consistent approach to the collection of assessment data for children with SSD. In order to evaluate different approaches to therapy and service models for children with SSD, we need to go beyond the collection of broad profession-wide measures such as the ROOT (RCSLT 2019) to collecting a minimal dataset which includes population specific pre- and post-intervention measures and relevant variables such as gender, age and a proxy for socio-economic status. This does not mean that outcome measures such as the ROOT should not be part of the minimal data set developed. While use of such outcome measures were not evidenced in this case note review, there is a clear need to consider the impact of the child's SSD, and their subsequent intervention, on activity, participation and well-being. The value of tools such as the ROOT and other impact-based outcome measures specific to SSD such as the FOCUS (Thomas-Stonell et al. 2010)and Speech Participation and Activity Assessment of Children (SPAA-C) (Mcleod 2004) could be investigated within a minimal dataset which also includes detailed measures specific to the population of interest (in this case children with SSD). The move to electronic health records could facilitate easy collection of such data. However transcription data is notoriously difficult to document electronically and systems allowing for this would need to be taken into consideration.

A consensus between clinical and research speech and language therapists regarding what should be collected for children with SSD is vital to ensure that the content of a future

minimal dataset is feasible to collect and fit for purpose. A minimal data set would need to feature a core outcomes set and other potential important variables. It would also need to be feasible to collect in clinical practice and provide a dataset for future investigations of clinically relevant research questions. Development work would be needed to determine the best way to establish such a minimal data set, however it may be beneficial to follow approaches already being carried out successfully to make minimal data sets with other clinical populations within SLT, for example those with children born with cleft palate (John et al. 2006, Wren et al. 2018a). This would need to involve a number of phases, with an initial review of the existing literature to generate preliminary content for existing outcomes and measures or instruments. It would also involve exploratory work establishing the range of interventions and subtypes of SSD and variables through consultations with clinicians. A final consensus would then need to be established, perhaps with the assistance of an expert panel, presenting a comprehensive list of outcome measurement and analysis instruments generated from the former phases. Qualitative research methods such as focus groups and Delphi methods that consult with these groups would likely be beneficial with an established process to work towards consensus.

Limitations

A limitation of the data used in this report is that data are from purposeful samples collected in clinical practice at three UK NHS sites. The samples are not necessarily representative of all NHS SLT services or broader practice, and comparisons between sites should be taken with this in mind. The data presented, demonstrate lack of assessment data in the instances of these three NHS sites and serves to support the discussion for the need for a minimal data set for children with SSD.

The informal discussion in this study with SLTs and service managers do not provide strong qualitative information and formal interviews with SLTs and managers may have further helped elucidate issues raised.

It should also be noted that the present study explores UK samples only. Although similar important issues may apply to international services, other issues to local contexts that have not been discussed here may apply.

Conclusion

Exploration of the case note data from three NHS sites indicate the need for a more consistent and contemporaneous collection of assessment data for children with SSD. The NHS has the potential to evaluate different interventions for children with SSD using data routinely collected in NHS SLT services but to achieve this we need to go beyond the collection of profession-wide measures such as the ROOT (RCSLT 2019) to establishing collection of a minimal dataset which includes population specific pre- and post-intervention measures and relevant variables. A consensus between clinical and research speech and language therapists regarding what should be collected for children with SSD is vital to ensure that the content of a future minimal dataset is feasible to collect and fit for purpose. The resulting dataset would be an invaluable resource to the clinical academic and research communities in addressing questions which have the potential to lead to improved outcomes and more cost-effective services.

Acknowledgments

This research was funded by Avon Primary Care Research Collaborative Research Capability Funding (Ref: 16/17 -3YW).

The authors would like to thank the speech and language therapy service leads and teams at the sites involved in data collection for the study

References

Almost, D. and Rosenbaum, P., 1998, Effectiveness of speech intervention for phonological disorders: A randomized controlled trial. *Developmental Medicine and Child Neurology*, 40, 319–325.

Armstrong, S. and Ainley, M., 1992;2012, *South Tyneside Assessment of Phonology*. Cornwall: Stass Publications.

Baker, E., McLeod, S., 2004, Evidence-based management of phonological impairment in children. *Child Language and Teaching Therapy*, 20, 3: 261-285.

Bates, S., Titterington, J. and the Child Speech Disorder Research Network, 2017, *Good practice guidelines for the analysis of children's speech samples in clinical practice and research*. Retrieved from:https://www.nbt.nhs.uk/sites/default/files/BSLTRU_Guidelines%20for%20Analysis%20of%20Child%20Speech%20Data%20FINAL%20v2%20Oct%202017.pdf

Bates, S. and Watson, J., 2012, *PPSA (Phonetic and Phonological Systems Analysis)* [online] Available at: http://sites.marjon.ac.uk/ppsa// [Accessed February 11th 2020].

Bowen, C., 1996, *The Quick Screener.* Retrieved on 1.04.2020 from www.speech-language-therapy.com

Bowen, A., Hesketh, A., Patchick, E., Young, A., Davies, L., Vail, A, Long, A.F., Watkins, C., Wilkinson, M., Pearl, G., Mo, Ralph, M. A. L. and Tyrrel, P., 2012. Effectiveness of enhanced communication therapy in the first four months after stroke for aphasia and dysarthria: a randomised controlled trial. *British Medical Journal*, *345*, e4407

Child Speech Disorder Research Network, 2017, *Good practice guidelines for the transcription of children's speech samples in clinical practice and research (2nd edition)*. Retrieved from http://development.nbt.nhs.uk/sites/default/files/Good%20practice%20guidelines_Transcription_2Ed_2017.pdf

Dodd, B., Zhu, H., Crosbie, S., Holm, A., and Ozanne, A., 2002, *Diagnostic evaluation of articulation and phonology (DEAP)*. Psychology Corporation.

Dobinson, C. and Wren, Y., 2019, Practice-based evidence: what is it and why do we need it? In Dobinson, C. and Wren, Y. (Eds) *Creating Practice Based Evidence: A Guide for SLTs*. Croyden: J&R Press.

Enderby, P. and John, A., 2015, *Therapy Outcome Measures for Rehabilitation Professionals*. 3rd edition. Guildford: J&R Press.

Fey M. E. and Finestack L. 2009, Research and development in child language intervention: A five phase model. In Schwartz R. (Ed.), *Handbook of child language disorders* (pp. 513–531). New York, NY: Psychology Press.

Gillon, G., 2000, The Efficacy of Phonological Awareness Intervention for Children With Spoken Language Impairment. *Language, speech, and hearing services in schools*, 31 2: 126-141.

Glogowska, M., Roulstone, S., Enderby, P., and Peters, T. J., 2000, Randomised controlled trial of community based speech and language therapy in preschool children. *British Medical Journal*, *321*(7266), 923.

Joffe, V. and Pring, T., 2008, Children with phonological problems: A survey of clinical practice. *International Journal of Language and Communication Disorders*, 43, 154-64.

John, A., Sell, D., Sweeney, T., Harding-Bell, A., and Williams, A., 2006, The Cleft Audit Protocol for Speech- Augmented: A Validated and Reliable Measure for Auditing Cleft Speech. *The Cleft Palate-Craniofacial Journal*, 43(3), 272–288.

Hegarty, N., Titterington, J. McLeod, S. and Taggart, L., 2018, Intervention for children with phonological impairment: Knowledge, practices and intervention intensity in the UK. *International Journal of Speech, Language and Communication Disorders,* 53(5), 995-1006.

Greenhalgh, T., 1997, *How to read a paper: The basics of evidence based medicine*. Chicago: *British Medical Journal Books*.

Hoffmann, T., Bennett, S. and Del Mar, C., 2010, *Evidence-Based Practice across the Health Professions*. Chatswood: Elsevier.

Kaipa, R., and Peterson, A. M., 2016, A systematic review of treatment intensity in speech disorders. *International Journal of Speech-Language Pathology*, 18(6) 507-520.

Kerryjane, M., Spilsby, K., 2006, *CLEAR: Phonology Screening Assessment*: Clear Resources.

Lee, A.S-Y., Gibbon, F.E., 2015, Non-speech oral motor treatment for children with developmental speech sound disorders. *Cochrane Database of Systematic Reviews* 2015, Issue 3, Art. No.: CD009383.

McCauley, R.J., Strand, E., Lof, G.L., Schooling, T., and Frymark, T., 2009, Evidence-based systematic review: Effects of non-speech oral motor exercises on speech. *American Journal of Speech-Language Pathology*, 4: 343-60.

McCartney, E., Boyle, J., Ellis, S., Bannatyne, S., Turnbull, M., 2011, Indirect language therapy for children with persistent language impairment in mainstream primary schools: outcomes from a cohort intervention. *International Journal of Language and Communication Disorders*, 46 (1), 74-82.

McLeod, S., 2004, Speech pathologists' application of the ICF to children with speech impairment. *Advances in Speech Language Pathology*, 6:1, 75-81.

McLeod, S., Harrison, L. J., and McCormack, J., 2012, *Intelligibility in Context Scale*. Bathurst, NSW, Australia: Charles Sturt University. Retrieved from http://www.csu.edu.au/research/multilingual-speech/ics. Published November 2012.

Moyse, K., Enderby, P., Chadd, K., Gadhok, K., Bedwell, M. and Guest, P., 2020, Outcome measurement in speech and language therapy: a digital journey. *BMJ Health Care Inform*, 27; e100085. doi:10.1136/ bmjhci-2019-100085

Murad, M. H., Asi, N., Alsawas, M., and Alahdab, F., 2016, New evidence pyramid. *BMJ Evidence-Based Medicine*, *21*(4), 125-127.

Murray, E., McCabe, P. and Ballard, K.J., 2015, A Randomized Control Trial of Treatments for Childhood Apraxia of Speech. *Journal of Speech, Language and Hearing Research,* 58 (3) 669-686.

Williams, P. and Stephens, H., 2004, Nuffield Centre Dyspraxia Programme 3rd Edition. *Nuffield hearing and Speech Centre*. The Miracle Factory: Windsor.

Palmer, R., Dimairo, M., Cooper, C., Enderby, P., Brady, M., Bowen, Latimer, N., Julios, S., Cross, E., Alshreef, A., Harrison, M., Bradley, E., Witts, H., and Chater, T., 2019, Self-managed, computerised speech and language therapy for patients with chronic aphasia post-stroke compared with usual care or attention control (Big CACTUS): a multicentre, single-blinded, randomised controlled trial. *The Lancet Neurology*, 18(9), 821-833.

Pring, T., 2004, Ask a silly question: two decades of troublesome trials. *International Journal of Language and Communication Disorders*, 39(3), 285-302.

Robey. R., 2004, A five-phase model for clinical-outcome research. .*Journal of Communication Disorders*, 37, 401–411.

Royal College of Speech and Language Therapists, 2019, *RCSLT Online Outcome Tool (ROOT)*, Different class solutions Ltd. Available: www.rcslt-root.org [Accessed 4 Nov 2019].

Ruscello, D.M., 2010, Collective findings neither support nor refute the use of oral motor exercises as a treatment for speech sound disorders. *Evidence-Based Communication Assessment and Intervention*, 4:2, 65-72.

Stevens,N., Isles, D., 2001, *Phonological Screening Assessment*. Bicester : Speechmark

Sugden, E., Baker, E., Munro, N., and Williams, A. L., 2016, Involvement of parents in intervention for childhood speech sound disorders: a review of the evidence. *International Journal of Language and Communication Disorders*, 51(6), 597-625.

Sugden, E., Lloyd, S., Lam, J., and Cleland, J., 2019, Systematic review of ultrasound visual biofeedback in intervention for speech sound disorders. *International Journal of Language and Communication Disorders*, 54(5), 705-728.

Taps, J., 2008, RTI Services for Children With Mild Articulation Needs: Four Years of Data. *Perspectives on School-Based Issues*, 9, 104-110.

Thomas-Stonell, N.L., Oddson, B., Robertson, B., Rosenbaum, P., 2010, Development of the FOCUS (Focus on the Outcomes of Communication Under Six), a communication outcome measure for preschool children. *Developmental Medicine and Child Neurology*, 52,47-53.

Wallace, S. J., Worrall, L., Rose, T., Le Dorze, G., Breitenstein, C., Hilari, K., Babbitt, E., Bose, A., Brady, M., Cherney, L., Copland, D. Cruice, M., Enderby, P., Hersh, D., Howe, T., Kelly, H., Kiran, S., Laska, A-C., Nicolas, M,m Patterson, J., Pearl, G., Rochon, E., Rose, M., Sage, K., Small, S., Webster, J., 2019, A core outcome set for aphasia treatment research: The ROMA consensus statement. *International Journal of Stroke*, *14*(2), 180-185.

Whiteneck, G. G., and Gassaway, J., 2013, SCIRehab uses practice-based evidence methodology to associate patient and treatment characteristics with outcomes. *Archives of Physical Medicine and Rehabilitation,* 94(4, Supplement), S67-S74.

Wren, Y.E., Humphries, K., Stock, N.M., Rumsey, N., Lewis, S., Davies, A., Bennett, R. and Sandy, S., 2018a, Setting Up a Cohort Study in Speech and Language Therapy: Lessons from The UK Cleft Collective Speech and Language Study. *International Journal of Language and Communication Disorders*, 53, 421-430.

Wren, Y., Harding, S., Goldbart, J., and Roulstone, S., 2018b, A Systematic Review and Classification of Interventions for Speech-Sound Disorder in Preschool Children. *International Journal of Language and Communication Disorders*, 53,446-467.

Tables and figures

Table 1 Case notes and blocks of therapy reviewed

Site	Number of case notes	Blocks of therapy	Which blocks of therapy
A	121	121	most recent block
B	12	31	all blocks r/v for each child
C	41	82	all blocks r/v for each child

Table 2 Service delivery agent, by site

	Site A (%)	Site B (%)	Site C (%)
Blocks delivered by speech and language therapist	15 (12)	11 (35)	80 (98)
Blocks delivered by SLT assistant	22(18)	15(48)	0 (0)
Blocks delivered by teaching assistant (+/- parent)	84(69)	5 (16)	2 (1.8)

Table 3 Transcribed assessment data available within one month of a therapy block

Site	Pre-intervention Assessment	Post-intervention Assessment
SITE A		
(n=121)	43.4%	32.7%
SITE B		
(n=31)	71.8%	21.8%
SITE C		
(n=82)	63.4%	50%

Table 4 Types of Assessments used (and percentage used overall)

Formal standardised assessment (%)	Published screening tools (%)	Informal sampling /none (%)
Diagnostic Evaluation of Articulation and Phonology (DEAP, Dodd et al. 2002) (1.2%)	CLEAR Phonology Screening Assessment(Kerryjane and Spilsby 2006) (16.7%)	In-house informal picture naming assessments (52.3%)
	South Tyneside Assessment of Phonology (STAP/STAP2, Armstrong and Ainley 1992; 2012) (3%)	Quick Screener (Bowen 1996)(0.4%)
	Phonology Screening assessment(PSA, Stevens and Isles 2001) (5.8%)	None (18.7%)
	Nuffield Centre Dyspraxia Programme (Williams and Stephens 2004) (1.7%)	

Figure 1 Assessment data collected at each site