Review

Emily Butler, Oliver Hugh and Jason Gardosi*

Evaluating the Growth Assessment Protocol for stillbirth prevention: progress and challenges

https://doi.org/10.1515/jpm-2022-0209
Accepted May 2, 2022; published online May 27, 2022

Abstract: Many stillbirths are associated with fetal growth restriction, and are hence potentially avoidable. The Growth Assessment Protocol (GAP) is a multidisciplinary program with an evidence based care pathway, training in risk assessment, fetal growth surveillance with customised charts and rolling audit. Antenatal detection of small for gestational age (SGA) has become an indicator of quality of care. Evaluation is essential to understand the impact of such a prevention program. Randomised trials will not be effective if they cannot ensure proper implementation before assessment. Observational studies have allowed realistic evaluation in practice, with other factors excluded that may have influenced the outcome. An award winning 10 year study of stillbirth data in England has been able to assess the effect of GAP in isolation, and found a strong, causal association with improved antenatal detection of SGA babies, and the sustained decline in national stillbirth rates. The challenge now is to apply this program more widely in low and middle income settings where the main global burden of stillbirth is, and to adapt it to local needs and resources.

Keywords: antenatal detection of SGA; DESIGN trial; fetal growth restriction; fetal growth velocity; Growth Assessment Protocol; small for gestational age (SGA); stillbirth.

Introduction

Stillbirth is a tragic event, even more so when it is avoidable. In high income countries, nine-tenth of intrauterine deaths occur in the antenatal period, with the largest category being fetuses that have failed to reach their growth potential [1]. Intrapartum related deaths may also have been preceded by fetal growth restriction, which is known to be associated with diminished fetal reserve for the stresses of labour.

Antenatal recognition of growth failure can halve the incidence of stillbirth [2]; therefore detection of fetal growth restriction (FGR) has increasingly been adopted as an indicator of the quality of maternity services, based on assessment of fetal size at delivery, i.e. small for gestational age (SGA) birthweight, as a proxy for FGR.

Fetal size needs to be monitored longitudinally to assess fetal growth, with protocols to tailor the frequency of assessment according to risk, as determined at the beginning of pregnancy and reviewed at each subsequent visits [3, 4].

Doppler flow investigation is an integral part of assessment and management to help determine severity of the problem and timing of delivery [5]. Their considerations are outside of this review, which focuses on a fetal growth surveillance programme and its role in stillbirth prevention.

Implementing the Growth Assessment Protocol

In many but not all health systems, antenatal care includes collaboration between primary (community), secondary (hospital) and tertiary (specialist referral) care, with a multidisciplinary effort including midwives, family practitioners, obstetricians, maternal-fetal medicine specialists and ultrasonographers, as well as consultation with neonatal services as required.

The aim in good maternity care is to make it as safe as possible while allowing the pregnancy to proceed as a natural process and reducing unnecessary interventions. It is with this philosophy that the Growth Assessment Protocol (GAP) was developed as a co-ordinated program for evidence based surveillance of fetal growth and prevention of stillbirth. It includes the following elements:

- A protocol and referral/care pathway which is regularly updated according to latest evidence and national guidelines, and adaptable to local circumstances and capabilities. The current GAP care pathway [6] incorporates principles espoused by the Royal College of Obstetricians
Evidence for customised charts

Underpinning effective fetal surveillance is a standard that can predict the expected birthweight and growth to reach it in a normal pregnancy, against which the actual growth can be measured. Unlike the usual separate fetal weight and birthweight standards in common use which do not align, GROW is a contiguous fetal and neonatal centile system which predicts normal pregnancy against which the actual growth can be measured. For example, version 2 of the SBL Care Bundle [8] advises against early term delivery based on SGA (3–10 centile) alone, because it increases the child’s risk of having special educational needs [9].

- Multidisciplinary training of all staff involved in antenatal care, in implementation of the care pathway, standardised measurement of fundal height, referral indications for ultrasound biometry to assess fetal weight and weight gain, and for investigation by Doppler where indicated.
- A baseline pre-implementation audit of SGA detection rate, followed by routine recording of the outcome of each pregnancy for ongoing monitoring of performance. In addition, audit of ‘near misses’ (SGA births which had not been identified antenatailly), using the GAP-SCORE electronic case review tool.

Customised standards are better than national or international population based standards in their association with stillbirth and neonatal mortality and morbidity [20–26] as well as placental pathology [27]. Further evidence comes from studies that show that standards based on populations with their own maternal characteristics are better associated with adverse outcome than an international one size fits all approach [24, 28]. The authors of the WHO fetal growth study [29], while producing a single standard based on data from eight countries, advised caution in its application as they observed in their own low risk cohort that significant variation in fetal growth exist due to variation in maternal characteristics in different populations.

Studies of subgroups categorised according to parity and maternal size for defining SGA birthweight [30] and fetal weight [31] showed that customising for these variables improves the association between SGA and stillbirth, and highlighted the high false positive and false negative rates of SGA defined by population based standards.

Adjustment for ethnic origin has been debated because of a belief that higher SGA rates in ethnic minority groups only represent social deprivation or malnutrition; however there is now substantial evidence that ethnic differences in birthweight are physiological and exist in low risk populations with pathological or socio-economic factors excluded [32–35]. Therefore, ethnicity needs to be adjusted for if the standard is to reflect the heterogeneity of the population; additional risk factors - which may occur more frequently in some ethnic groups - need to be recognised and taken into consideration when planning care – for any ethnic group.

It is clear that one size does not fit all [36]. But despite all the direct and indirect evidence produced since the concept of customised charts was first published 30 years ago [37], there is still controversy in some quarters, which has been presented and responded to in a recent expert review [38]. For clinicians who are still unconvinced about whether to adjust for one or other of the constitutional factors, or if they do not have a record of such information, the GROW chart software can also produce partially adjusted curves, using population average values for the variables not specified; for ethnicity a ‘global average’ can be used, based on the GROW database of 4 million births with over 100 ethnic origin categories.

However when producing the chart, the more information, the better: the predictive power of the model increases with each variable added, reaching an R square...
value of 0.76 for predicting normal birthweight (defined as within the mid tertile of the birthweight distribution of the population; Figure 1).

Customisation for constitutional maternal characteristics improves the distinction between physiological and pathological smallness, and reduces false positive diagnoses of ‘SGA’ which can result in unnecessary investigations and intervention [39, 40]. Abnormal measurements are more likely to be taken as a serious warning rather than ‘normalised’ and ignored, clinicians’ confidence in their assessment of growth is improved and they are more ready to act, as and when required.

The value of customised assessment in case audit and research has been established through the above mentioned studies of databases large enough to show the strength of association with adverse outcome. They are provided as free tools for clinicians and researchers to produce customised fetal and birthweight centiles for individual pregnancies (GROW ICC – individual centile calculator; currently used by 270 clinicians in 61 countries) or for calculating centiles in databases (GROW BCC – bulk centile calculator; 175 researchers in 46 countries) [41].

For clinical application, GROW charts and centile calculators constitute an essential part of the Growth Assessment Protocol (GAP) for stillbirth prevention. The GAP program includes training, GROW software, audit tools, automated and bespoke reporting and helpdesk support. It is provided by the Perinatal Institute, a not-for-profit social enterprise, under license for nominal cost.

Evaluation of GAP

Acceptability and compliance

Clinical effectiveness can be evaluated at unit, regional and national levels. At unit level, the first task is to make the program acceptable and implementable as part of routine care. Training is paramount, and often there are insufficient resources, in particular in the availability of ultrasound scans for suspected growth restriction. The protocol needs to be aspirational, while flexible for local adaptation to allow for limitations in capacity. Being a program requiring multidisciplinary change management, GAP usually takes 6–12 months or even longer to properly ‘embed’ in the clinical service.

GAP’s acceptability is evident by its wide uptake in the UK National Health Service (NHS), where it is currently operational in 113 (74%) of all NHS hospitals [42] with about 450,000 births per annum. While training is a prerequisite, there are different levels of implementation, categorised mainly according to whether the unit engages in audit by routinely recording its birth outcomes, from which it can derive its own SGA detection rate to monitor and benchmark performance and identify their near misses for case review.

Of the 113 GAP units in the NHS, 97 (86%) are ‘complete’ implementers and record the required information in over 90% of their births. At parents’ request, the list of GAP units is published and regularly updated on the Perinatal Institute’s website, together with the level of implementation (partial or complete) [42].
Detection of SGA

Stillbirth numbers at unit level are insufficient for trends to become easily apparent. Hence antenatal detection of SGA is used as a proxy indicator, with SGA at birth as denominator. Before implementation, all units are asked to review a random sample of pre-GAP cases – about 10% of their annual deliveries – to determine their baseline pre-GAP detection rate. Following implementation, a core dataset of birth details including antenatal referral and detection are recorded after each delivery. This then becomes part of the recommended rolling audit to monitor performance, and can be used by local units and regional or wider networks for benchmarking. The Power-BI reporting function of the GROW App displays for each unit its own performance, together with the national GAP average and the ‘Top 10’ best performing units, providing a benchmark and an indication of what can be achieved by units working in the same health system (Figure 2).

From this unit based information, anonymised data are compiled to ascertain national trends. This is displayed in Figure 3, with information from the 97 NHS hospitals submitting routinely recorded data.

Figure 2: Power BI report.
Example of a recent quarterly report for a GAP unit in the UK national health service, with detection rate in the national ‘top 10’. False positive detection rates are also shown.

Figure 3: Trend in antenatal detection of SGA (<10th centile). GAP average (GAV), Top Ten and false positive (FP). Error bars = 95% confidence interval. Q1 2020/21 dip was associated with national COVID-19 lockdown.
Baseline SGA detection rate pre GAP implementation averaged 19%. After training and local adaptation and implementation of the protocol, average detection rates gradually increased twofold, running at just above 40%. Limited ultrasound resources are often quoted as a reason why this average is not climbing higher. However ‘Top 10’ detection rate of SGA <10 centile has increased more than threefold, to 63.1%, indicating that such rates are achievable. For SGA <3rd centile at birth, the average antenatal detection rate is 61% overall and 82% in Top 10 units.

Local GAP-SCORE reviews of missed cases have helped clinicians to highlight system failures and to make the case for better funding for maternity services including staffing and ultrasound resources.

Improvements in SGA detection rate as the main outcome indicator have also been shown in studies of standardised fundal height measurement plotted on customised charts [43, 44] and in studies evaluating implementation of the GAP program in units in Australia [45] and New Zealand [46].

Effect on stillbirth rate

Assessing effect on stillbirth rate requires a longer time frame or wider geographical area. Analysis of regional data from the UK Office of National Statistics (ONS) following implementation of GAP in three NHS health regions in England showed a significant decline in stillbirths, while there was no change over the same period in the rest of the country [47]. There were no identifiable changes in clinical practice or in health services in the three study regions apart from the implementation of GAP, and examination of all causality criteria demonstrated that, although the study was observational, the association was causal, i.e. the reduction in stillbirths was attributable to GAP [47].

A report based on data from Scotland claimed that stillbirths there declined already before GAP was implemented [48]. However the study was demonstrably flawed [49]: although the information was available well before their analysis was undertaken, the authors failed to include data from the following years where the stillbirth rate rose again, showing that the pre GAP drop was temporary and probably associated with a concurrent research study with higher scanning rates being rolled out in Scottish units [50]. Following subsequent GAP implementation, the stillbirth rate showed a sustained decline, in line with trends observed in English regions [49]. This improvement was consistent also with unit based findings in a separate Scottish study which reported that in 2014/15, SGA detection rates were found to have improved only in the three units that had implemented GAP by then [51].

In England, an independent evaluation [52] was commissioned by NHS England to evaluate the introduction in 2016 of the Saving Babies’ Lives Care Bundle [7], which included an element for fetal growth assessment modelled on the GAP algorithm and referral pathway [4]. The study observed a reduction in stillbirths and an associated increase in antenatal detection of SGA fetuses. Almost all (15/17; 88%) of the hospitals submitting data for this analysis were in the GAP program.

Comparing against national trends

An award winning study [53] of ONS stillbirth data from England looked at the 10 year trend from 2008 (pre GAP) to 2017. Maternity units were analysed in three groups according to status of GAP implementation in 2017: none, partial and complete. These categories are displayed and updated online [42]. ‘Complete’ was defined as detailed above – i.e. units that routinely recorded detection rates and birth outcome, and reviewed their missed cases. A fourth category was a subgroup of the 20 best performing units according to SGA detection rates in the ‘complete’ group.

We found that in 2008, each of these categories of units had similar stillbirth rates (Figure 4). In 2017, the histogram shows 1. an overall drop in stillbirths since 2008 in all groups, including in the non-GAP units; 2. a similar drop in the ‘partial GAP implementer’ units, 3. a significant decline in the ‘complete GAP implementer’ units; and 4. an even steeper drop in units with the best antenatal detection rates. These findings suggest a dose dependent relationship: the better the fetal growth surveillance, the steeper the decline in stillbirths.

The 10 year trend in the non-GAP sites [53] suggested the start of a year-on-year decline in stillbirths from around 2013, coincidental to the publication of the 2013 RCOG SGA guideline [3]. The guideline helped raise awareness of the importance of early risk assessment and appropriate investigation and management according to risk. The importance of serial scanning in high risk pregnancies is supported by recent evidence from GAP units that pregnancies in which increased fetal risk is identified but then not monitored with serial biometry have a substantially higher rate of stillbirth [54].

Randomised controlled trial (RCT): the DESiGN study

RCTs are rightly considered the gold standard for evaluating alternative forms of treatment. However their
The applicability to evaluate a surveillance program aimed at antenatal detection of SGA, which does not prescribe the actual management once SGA is detected, is debatable. Also, because of a strong element of training and change of protocol/practice, randomisation cannot be case by case but has to be by units or ‘clusters’.

Such a cluster RCT, the DESiGN Trial [55], was set up in 2015/6 by a group of investigators in London to assess the effectiveness of the GAP programme, with the primary outcome measure being antenatal detection of SGA. The Perinatal Institute’s GAP team agreed to support the trial and provide training, GROW charts and audit tools for units that had been randomised to the study group. Assurances were given by the investigators that they will ensure that the program is fully implemented at unit level according to a set of GAP criteria agreed in 2015, prior to commencement [56]. The trial ran from 2017 until data collection was stopped in February 2019. At the time of writing (April 2022) the findings of the study have not yet been published.

The GAP team produced their own internal report [57] based on their observations while providing training and helpdesk support, and from the data routinely recorded in the GROW App by clinicians in the study arm. The findings were also communicated to the chief investigator and the chair of the steering committee of the DESiGN Trial.

The report raised a number of concerns about methodology and manner of execution of the trial which can be summarised in the following points:

1. Inadequate control group: during the period of the trial, NHS England’s national roll-out of its Saving Babies Lives Care Bundle [7] was – despite previous assurances – also allowed to include the control arm units (clusters) of the trial. The fetal growth element of the Bundle had a surveillance algorithm ([7], p19) that – apart from not specifying which charts to use – was a copy of the one developed by the Perinatal Institute in 2013 for its GAP program [4] and which was reproduced in the DESiGN protocol since its early (2015) versions ([56], p30).

2. Lack of priority: Most units in London had already opted to be in the GAP program before the trial, or dropped out before randomisation as they wanted to proceed with immediate implementation. For the remaining units that were subsequently randomised into the study arm, implementation seemed low priority and had to compete with other priorities. One result of this was that there were delays of up to 12 months between staff being trained and the unit going live, with much of the hands-on training content likely to have been forgotten by then, and little or no take up of the GAP e-learning module supplied. Furthermore, implementation of the agreed minimum criteria [56] failed to be achieved in all participating units.

Figure 4: ONS stillbirth rates 2008 (pre GAP) and 2017, in non-GAP and GAP units categorised according to partial and complete implementation, and a ‘top 20’ subgroup with the highest detection rates.
(3) Trial period too short: the delays were not taken into consideration by the investigators and data collection was stopped in February 2019. As a result, in five of the six units implementation was incomplete during the period of the study, and in the 6th unit, it was achieved only in early 2019, as detailed in the GAP team’s report [49].

According to data entered into the GROW App at unit level, three of the units/clusters did achieve complete implementation in terms of routine recording of birth data, but only during 2019. Analysis of their SGA detection shows significant improvement, reaching rates similar to the national GAP average, but only after the February 2019 end of the trial’s data collection period (Figure 5).

The trial will therefore not be able to provide valid answers to the question of how GAP performs in units that are interested in implementing GAP in their practice. It will instead highlight some of the challenges in running an RCT, including the need to find units that are willing as well as able to participate; to allow sufficient time for a complex intervention to ‘embed’ in clinical practice before it is evaluated; and to ensure that the RCT’s control group continues to provide the same ‘standard care’ during the trial as it did before. The longitudinal, observational studies referred to earlier are more likely to provide realistic and externally valid assessments.

The DESIGN study will also not be able to evaluate customised vs. population based growth and birthweight standards, as this requires larger, epidemiological datasets with ‘hard’ outcome, as listed above. Prospective clinical trials of different surveillance methods need to agree on one standard to use. According to their protocol [55], the investigators decided to define outcome as ‘SGA by both methods’, these being GROW customised birthweight centiles and the UK90 (UK-WHO) neonatal standard [58]. However ‘SGA by both methods’ does not take into account that the customised standard is more sensitive in detecting SGA cases that represent increased risk. A comparative study in 2015 [59] of a regional dataset of over 140,000 births showed that 30% additional cases were SGA by GROW only: these pregnancies had a significantly increased risk of adverse outcome including stillbirth (OR 3.6, CI 2.8–4.7) as well as neonatal deaths, low Apgar scores and admission to neonatal intensive care [51]. In contrast, 19% of cases were SGA by UK90 only, and these had no significant associations with any of the adverse outcome measures studied. As GROW is a contiguous fetal and neonatal standard, the additional at-risk cases would also have been identified during antenatal care.

Further work

Assessment of SGA detection rates in clinical care is important as an auditable indicator of performance, in identifying cases with elevated stillbirth risk [2]. However it needs to be complemented by identification of slow growth, which has long been known to be also associated with adverse outcome [60, 61]. Recent studies have reported on the association between slow growth and adverse outcome in the absence of SGA [62–65]. Varying definitions for ‘slow growth’ have been proposed, such as a drop by 30 centiles [63] or two quartiles [66], but without specifying over what interval. Another suggestion was to use a lower limit of 20 g
per day [8] based on a study which however itself acknowledged that normal growth rate varies with normal constitutional characteristics affecting fetal growth [67]. We analysed a large database of serially scanned pregnancies and defined measurement interval specific cut-offs for normal growth [68]. The method was evaluated in terms of its association with outcome and showed firstly that the new limits for normal growth rate/velocity predicted normal outcome, while slow growth was significantly associated with stillbirth. Importantly, in 66% of pregnancies with slow growth and subsequent stillbirth, the last scan-estimated fetal weight did not indicate SGA.

We have incorporated automatic recording and display of growth velocity into the new, electronic version of the customised growth chart (GROW 2.0). Figure 6 shows a screenshot with an example of slow growth. It is hoped that this functionality will help identify additional pregnancies at risk – for further investigation, timely management and prevention of adverse outcome.

**Conclusions and next steps**

Fetal growth restriction is not the only condition associated with stillbirth, but the most frequent and most avoidable. Evaluation of a prevention program focussing on antenatal detection of the at-risk baby has shown it to be acceptable and effective. It is also inexpensive, as it titrates the level of surveillance to the prevalent risk, and is hence suitable for adaptation and roll-out in low and middle income settings, where the main global burden of stillbirth is. Encouraging early signs in South Asia and Africa are showing that implementation is feasible, with local champions driving improvements in quality and safety of maternity care.

**Research funding:** None declared.

**Author contributions:** All authors participated in conceptualising this review article and in the write up of the manuscript, accept responsibility for its content and approve its submission.

**Competing interests:** All authors work for the Perinatal Institute, a not-for-profit social enterprise which provides the GAP program and associated training and software licenses as a service.

**Informed consent:** Not applicable.

**Ethical approval:** Not applicable – all data were fully anonymized.

**Data availability:** The datasets described in this review are available on reasonable request, but restrictions may apply for data collected under license from hospitals in the National Health Service.
References


